It Takes Two Hands to Clap:

Effects of Reputation and Search in Healthcare Markets[*]

Ming Ge[a*]

[a] University of Massachusetts, Amherst, United States

[*] **Corresponding author**

Author Name Ming Ge

Postal address: 80 Campus Center Way, Amherst, MA 01002

E-mail: mge@umass.edu

[*] I thank Christian Rojas and Rong Rong for their invaluable guidance and advice. I also thank Juan-Camilo Cardenas, John Spraggon, and Xiaolu Wang, as well as conference participants at ESA 2022, New England Experimental Economics Workshop (NEEEW) 2022, and UMass ResEcon Graduate Conference 2021 for thoughtful comments and feedback. Financial support from the Graduate School and the Department of Resource Economics at the University of Massachusetts Amherst is gratefully acknowledged.

Abstract

Asymmetric information between physicians and patients often leads to rampant over-treatment and low market efficiency. A standard reputation system falls short of being effective given the credence good nature of this market: patients cannot tell whether a high-cost treatment recommendation (versus a less costly and complex treatment) is necessary even after the service is completed, which creates substantial incentives for physicians to overtreat. This study proposes a novel solution to reinstate the function of reputation by combining a reputation mechanism with patient search for second opinions. The key insight is that patient search can help keep physicians' honesty in check, thereby facilitating reputation-building in a repeated interaction setting. This mechanism was tested in a laboratory experiment, and its effectiveness in reducing overtreatment was confirmed. Additionally, the study manipulates search costs and examines their impact on overtreatment. The findings show that search cost matters as it limits the role of patient search and, consequently, the effective size of physician reputation is smaller in the high search cost condition. The results highlight that accessibility to patient searches may contribute to more equitable access to healthcare.

Keywords: Reputation, Overtreatment, Search Costs, Healthcare Market, Credence Goods Market

JEL: C90, D82, D83, I10, L14, L84

1. Introduction

Healthcare, a prominent example of a credence good[1], is characterized by several factors that dampen market efficiency. Early work by Darby and Karni (1973) highlighted that a fundamental reason for such inefficiencies is the lack of information symmetry in credence goods markets; patients are less informed than physicians regarding the optimal treatment for their health problems. Therefore, uninformed patients must rely on physicians to diagnose their problems and provide appropriate treatment recommendations. While patients can observe the treatment outcome (e.g., their recovery from illness), they may never know whether a less expensive treatment could have achieved the same results. This leaves little room for patients to accumulate and share information about their experiences regarding physicians' conduct. Patients' lack of ability to identify dishonesty naturally leads to physicians' opportunistic behaviors (e.g., prescribing procedures, tests, or medications that are more than necessary to treat the disease), commonly known as the "overtreatment problem." Overtreating a patient is usually associated with a greater marginal profit for healthcare providers and higher medical expenses for patients[2]. In addition to the excess payments made from patients to providers, valuable resources are also wasted to conduct these unnecessary treatments, which leads to an overall loss of efficiency in healthcare markets.

According to an American Medical Association survey, in the U.S. healthcare market alone, an average of 20.6% of medical care was unnecessary, including 22.0% of prescription medications, 24.9% of tests, and 11.1% of procedures (Lyu et al., 2017). A salient example of overtreatment can be found in Gruber and Owings (1996) and Johnsen and Rehavi (2016). Their studies suggest that OB/GYNs prefer cesarean delivery, an overtreatment with a high reimbursement rate, over normal childbirth. In another instance, evidence from prescription drugs in China and Japan shows that physicians tend to prescribe drugs with higher markups (Iizuka, 2007; Currie et al., 2014). More broadly, evidence of overtreatment in a variety of medical settings can be found in Domenighetti et al. (1993), Delattre and Dormont (2003), Brownlee (2010), and Gottschalk et al. (2020). In all these cases, physicians'

---

1 In addition to medical services, many other markets for professional services, including repair services (Schneider 2012), taxi rides (Balafoutas et al. 2013), and management consulting (Craig, D 2005), also exhibit properties of credence goods and suffer from market inefficiencies caused by information asymmetries.

2 Overtreatment is also driven by physicians' risk aversion toward malpractice liability. In this study, I focus on physicians' overtreatment fueled by financial incentives. Thus, I assume physicians can perfectly diagnose the problems at no cost.

overtreatment not only leaves more costly medical bills to patients[3] but also generates large amounts of wasted resources as more complex tests and treatments are typically costlier. In 2010, the Institute of Medicine called attention to the problem, suggesting that "unnecessary services" are the largest contributor to waste in U.S. healthcare markets (McGinnis et al., 2013).

There is an intuitive and widely accepted solution to mitigate physicians' overtreatment behavior, allowing patients to search for a second opinion. That is, patients are permitted to elicit multiple recommendations from different physicians and choose the one that is most cost-effective in treating their problem.[4] Health insurers and government legislation in the U.S. and many European countries recommend this strategy (Hu, 2017; Pieper et al., 2017). There are two potential reasons why patient search can alleviate overtreatment and generate higher market efficiency. First, patients' ability to search can increase the intensity of competition among physicians. When a physician recommends expensive treatment, the patient is more likely to visit a different physician. This mechanism should have its intended benefits, whether the patient seeks treatment only once (what we refer to as "one-shot" interactions) or in cases when patients might seek treatment multiple times (what we refer to as "repeated market interactions"). A key difference between one-shot and repeated interactions is that a physician's reputation building does not exist (or matter) in the former but does in the latter. Importantly, reputation building can significantly increase the effectiveness of patient search for reducing overtreatment (and, therefore, increasing market efficiency even further). Patients may avoid future visits in repeated interactions once a dishonest physician is identified. This gives physicians an additional incentive to refrain from overtreating so that they can avoid the risk of losing future business.

As the real-world healthcare market always involves repeated interactions, both competition intensity within a market period (facilitated by the role of patient search) and physicians' reputation building (facilitated by patients' ability to report on or access healthcare providers' prior treatment recommendations) can be important factors in determining a market's level of observed overtreatment. This makes a laboratory experiment an ideal tool for identifying and quantifying the effectiveness of

[3] Brownlee (2011) notes that overtreatment also makes medical errors more likely, because the higher the volume of care you receive, the greater the odds are that somebody, somewhere, will make a mistake. As this study concentrates on the cost inefficiency of overtreatment, I assume that there are no adverse health effects from being overtreated.

[4] Another advantage of searching for multiple opinions is to avoid incorrect diagnoses. In this study, we focus on its merit of restraining overtreatment only by excluding the possibility of misdiagnosis from the physician's choice set.

reputation by comparing its combined effects with patient search with the effect of patient search alone. To achieve this goal, this study designs a multi-seller-multi-buyer credence goods market in which informed sellers have the incentive to overtreat and uninformed buyers can always search for an additional treatment recommendation from a different seller. Under the baseline condition, I turn off the possibility of a reputation mechanism by randomly reshuffling seller IDs after each market period. The random ID design effectively turns a repeated game into many one-shot interactions where reputation-building is impossible[5]. This allows me to measure the effectiveness of the market competition mechanism (i.e., as facilitated by second opinions) alone in reducing overtreatment. In the "reputation" condition, I fix the sellers' IDs and make the history of buyers' past transactions exclusively visible to them[6]. Through this, a seller can build a reputation (or risk ruining it) by choosing proper treatment (or overtreatment) for buyers. Note that the aforementioned market competition mechanism also exists in this treatment. The study experiments with general credence goods framing to avoid a framing effect that may confound the main results. For the purpose of this study, I will refer to sellers as physicians and buyers as patients.

Notably, unlike in experienced goods markets, where information asymmetry disappears after consumption, reputation only functions properly in a credence goods context with buyer search. This is because, for credence goods, the detection of overtreatment only occurs through a mismatch of treatment recommendations from different sellers. If the search opportunity is removed from the market, buyers will never be able to detect overtreatment. This renders information regarding past transactions useless. Dulleck et al. (2011) have experimentally investigated the role of reputation for credence goods in the absence of a buyer search. They find that reputation alone does not reduce overtreatment and inefficiency in the credence goods market. Mimra et al. (2016a) also experiment with multiple formats of reputation systems without allowing buyers to search. They confirm that varying information content in the reputation system does not affect sellers' overtreatment behavior. This is not surprising since buyers in these markets can only observe a series of past recommendations with no reputation value, and an honest high-cost treatment recommendation can never be

---

[5] While it is possible to run one shot games with different sets of buyers and sellers each time, the cost of such design could be prohibitive.

[6] This is different from displaying past history of all market transactions, which is harder to achieve in healthcare markets for privacy reasons.

distinguished from an overtreatment. This study reconciles these null results and proposes a functional reputation-building system by ensuring that buyer search is an essential part of market design.

While buyer search improves market efficiency by reducing costly overtreatment, it comes with an obvious caveat that searching for additional opinions is costly[7]. In healthcare markets, the fees for second opinions are not negligible and vary significantly from patient to patient. If the search cost is sufficiently high, duplication of diagnoses could entirely offset the efficiency gain from reducing physician overtreatment.[8] This makes the level of search costs in the market a critically important parameter. Wolinsky (1993) and Mimra et al. (2016b) theoretically show that search costs could be prohibitively high to prevent buyer searches and limit the improvement of market efficiency.

The level of search costs is critical to my investigation of the effect of reputation building: when patient search is limited by high search costs, it becomes less practicable for patients to recognize who is a dishonest physician. Therefore, physicians have no incentive to build a reputation. In other words, a high search cost practically converts a market with patient search and reputation discovery into one without them. I hypothesize that high search costs reduce and eventually eliminate the reputation effect. To empirically test this hypothesis, I introduce another treatment dimension by varying the cost for patients to search for a second recommendation. Under high search cost conditions, the impact of reputation on overtreatment is again measured. I find that compared to the low search cost condition, the reputation effect is significantly reduced under a high search cost. These results are important for the current discussion on healthcare inequality, as patients in lower-income brackets and with poor insurance coverage may have higher search costs. This would prevent them from receiving full benefits in terms of cost-effective medical treatment owing to prohibitive search costs.

The main conclusion of my study is that reputation is an important mechanism for the effectiveness of patient search in reducing overtreatment. The average overtreatment level decreases from 38.54% to 21.67% when reputation is at play. However, this reduction is only detectable in the low search cost treatment, which confirms the hypothesis that search costs need to be sufficiently low for overtreatment

---

[7] Wagner and Wagner (1999) report that of persons who had visited a doctor in 1994, 19% received a second opinion. The cost associated to searching for a second opinion is estimated at $3.2 billion. This is likely an underestimate as it does not factor in opportunity costs.

[8] The costs spent on searching for a second opinion include, but are not limited to, a second diagnosis cost, transportation costs and time costs. I refer to these costs collectively as search costs.

detection to be sizeable and, therefore, for reputation building to have a bite. In addition, it was observed that patients search less frequently when physicians build a reputation. The reason for this is that patients receive low-cost treatment recommendations more often on their first visit, which in turn reduces their need to search. As a result of lower overtreatment and fewer searches, market efficiency significantly rises when reputation exists.

This study contributes to the literature by experimentally analyzing how reputation impacts overtreatment and market efficiency in healthcare markets. Contrary to all the earlier studies, which found a null effect of reputation in credence goods markets, my experiment highlights that patient search is necessary to establish a meaningful reputation mechanism in such markets. To the best of my knowledge, this study is the first to quantify the positive effect of reputation on reducing wasteful overtreatment in an experimental healthcare market. The comparison between low and high search costs in this environment sheds light on a previously unexplored channel where the difference in search costs (e.g., insurance coverage, health care budget) can further exacerbate the problem of health inequality.

The remainder of this paper is organized as follows. Section 2 reviews related literature. Section 3 presents the experimental design of the study. Section 4 provides the hypotheses. Subsequently, Section 5 reports the experimental results. Finally, Section 6 concludes the paper and discusses the policy implications.

2. Literature Review

A large and growing body of literature has studied the different mechanisms that discipline seller behavior in credence goods markets.[9] Two related strands of literature address the overtreatment problem in such markets from different angles. The first group of studies examines the effect of buyer search in a one-shot market, while the second focuses on the formation of reputation in repeated interactions.

Several studies have analyzed the impact of buyer search in a one-shot interaction. Wolinsky (1993) provides an earlier theoretical model describing the incentive to overtreat in credence goods markets.

[9] Dulleck and Kerschbamer (2006), Kerschbamer and Sutter (2017), and Balafoutas and Kerschbamer (2020) provide comprehensive reviews of the literature on when and how credence goods can be provided efficiently.

His model demonstrates that introducing costly buyer searches can prevent sellers from overtreating. Sellers in this model choose not only the treatment recommendations but also their corresponding prices. One of the model's predictions is a symmetric equilibrium, which is consistent with the experimental results of this study, where the amount of overtreatment is determined by the level of search cost[10]. Mimra et al. (2016b) modify Wolinsky's model by focusing on cases in which the treatment prices are predetermined. They conduct the first laboratory study to examine the impact of costly buyer searches on the rate of overtreatment and market efficiency in a one-shot market environment. They find that when the search cost is sufficiently low, introducing the possibility of searching for a second opinion significantly reduces the level of overtreatment, thereby improving market efficiency (high search costs significantly dampen this effect). This result is replicated by Agarwal et al. (2019) in a slightly different laboratory setting. In Agarwal et al. (2019), buyer search only acts as an information source, and buyers cannot undergo treatment with second-selected sellers. It is worth emphasizing that the market designs of Mimra et al. (2016b) and Agarwal et al. (2019) use a random matching procedure between market periods. The absence of buyers' choice of which seller to visit removes the incentive for sellers to build a reputation for future businesses. This study complements their work by examining how buyer searches can improve credence goods market efficiency through the critical reputation channel.

A plethora of research suggests that reputation is crucial in facilitating trade in markets with asymmetric information. Most of these studies look at experience goods markets[11], where buyers know exactly what they need ex-ante and learn the value of the goods or services through consumption. In the case of experienced goods markets, past sales provide an objective metric for gauging a seller's product quality, thereby reducing fraud. This differs from credence goods, where buyers cannot detect overtreatment via consumption. This sharp distinction between experience goods and credence goods is captured in the results of the laboratory and field experiments. Reputation has been shown to reduce moral hazard and improve the market outcomes for experience goods (Dellarocas, 2006; Huck et al., 2012; Tadelis, 2016). However, in the context of credence goods, reputation alone appears to be

[10] An additional asymmetric equilibrium exists where some sellers specialize in being the ones who conduct only low-cost procedures, while the rest of the sellers can conduct either high-cost or low-cost procedures.

[11] See Bar-Isaac and Tadelis (2008) for a literature review on reputation and trust in experience goods markets.

ineffective in limiting sellers' overtreatment or improving market efficiency (Dulleck et al., 2011; Schneider, 2012; Mimra et al., 2016a[12]).[13] This study highlights the unique characteristics of credence goods and explores the potential of reputation in the more fitting institutional environment of such goods.

Two recent theoretical studies are closely related to this study. Fong et al. (2022) model an infinitely repeated game in which long-lived sellers interact with short-lived buyers who can verify whether the selected seller makes an unnecessary treatment recommendation by searching for a costly second opinion. A seller loses all future businesses if his overtreatment is detected by a buyer. Moreover, new buyers observe the past transaction history of all buyer-seller pairs. As both design features are unlikely to be held in actual healthcare markets, this study considers a more realistic setup by introducing two major changes. First, punishment for dishonest physicians was endogenously determined by the patients in my experiment. Second, the experimental design allows patients access to past transactions and search history involving themselves (and not other patients). This feature is consistent with the fact that other patients' experiences with a physician are protected by HIPAA authorization and are unlikely to be publicly available.

Gerlach and Li (2022) theoretically and experimentally study the level of overtreatment in both monopoly and duopoly markets. They also investigate two different scenarios in a duopoly market, with and without buyer search. They do not detect any significant impact of introducing buyer search from their experimental results, despite the efficiency-enhancing effect predicted by theory. A critical market feature that may contribute to this result is that buyers in all markets are allowed to purchase any treatment independent of the seller's recommendation. This "freedom of choice" design enables buyers to detect overtreatment whenever a high-cost recommendation is followed by a buyer's successful attempt to disregard such a recommendation. In other words, their experiment points out that buyer search can be redundant when freedom of choice is possible. Results from Gerlach and Li (2022) apply directly to credence goods such as business consulting and repair services where

[12] In their set-up, reputation-building incentivizes sellers to provide their buyers with sufficient services, but it could not restrain sellers from overtreating their buyers.

[13] The results are contradictory to the theoretical predictions in Wolinsky (1993) and Frankel and Schwarz (2009). In these two theoretical papers, the authors demonstrate that sellers may have the proper incentives not to overtreat their buyers if buyers give more business to sellers who previously provided low-cost treatments.

"freedom of choice" for buyers is common. However, in healthcare markets, patients can rarely undergo procedures or tests that are not recommended by a physician. This study's experimental design provides a more suitable environment for studying the effect of reputation in the healthcare domain.

3. Experimental Design

3.1 Market Procedure

This study constructs a basic market structure with exogenous prices,[14] following the design of Mimra et al. (2016b)[15]. Each market period consists of four patients and four physicians[16]. At the beginning of each session, the subjects are randomly assigned to be either a patient or a physician. Assignments are fixed for the entire experiment. Each experimental session has 20 market periods.

The decision sequence for patients and physicians in a particular market period is as follows. First, each patient independently receives a randomly determined type of problem. The problem can be either a major problem with probability $H=0.25$ or a minor problem with probability ($1\text{-}H=0.75$). Patients have no information on the type of problem they have. Thus, in the next step, each patient must select one of the four physicians to provide a treatment recommendation. Simultaneously, each physician is asked to provide a treatment recommendation for each patient after observing the type of problem they have[17]. The treatment can be either low-cost or high-cost. A high-cost treatment can fix major and minor problems, but a low-cost treatment can only fix minor ones. Note that physicians are liable to cure their patients;[18] in our setting, physicians cannot provide low-cost treatment recommendations to patients with major problems. After a patient selects a physician, he observes the treatment recommendations offered by the physician. Consequently, the patient has two options: (1) accept the treatment recommendation and let the selected physician perform the corresponding treatment, or (2)

---

[14] Fixed prices are common in the U.S. healthcare market where prices are set as a result of a centralized bargaining process (Sülzle and Wambach 2005).

[15] Equilibria under our baseline conditions (no reputation) are characterized in their study.

[16] I decide to employ four participants per role because collusions are rare in markets with four or more sellers (Brandts and Potters 2018).

[17] We implement the strategy method for three reasons: (1) so that physicians do not know whether a patient is on her first or second search, a condition required by the model in Mimra et al. (2016b); (2) it collects more decisions from physicians per period.

[18] This assumption is valid in healthcare markets because leaving patients uncured not only violates the Hippocratic Oath and the World Medical Association's Declaration of Geneva, but also incurs a medical malpractice lawsuit.

pay a search cost K to select another physician from the three remaining physicians and undergo the treatment performed by the second selected physician.[19]

The payoff function π for each patient j is defined as the difference between a fixed value V=130 points, utility of recovering from the unknown problem, and price of the recommended treatment Pt (P_H=115 points; P_L=75 points). Clearly, patients prefer low-cost treatment recommendations to high-cost ones. Moreover, if patients choose to search for a second opinion, they pay an additional search cost of K.

$$\pi_j = \begin{cases} V - P_t & \textit{if the patient j accepts} \\ V - P_t - K & \textit{if the patient j searches for a second recommendation} \end{cases}$$

Another important feature of the above patients' payoff function is that if a patient does not conduct a second search, he/she cannot detect whether the physician is honest (by way of observing the physician's payoff). However, if a patient chooses to search for a second recommendation, discovering a physician's dishonesty becomes possible, although not guaranteed. Detection only occurs when the first recommendation is high-cost, whereas the second recommendation is low-cost.

Physician i receives payoff e which is calculated as the difference between the price and cost of the accepted treatment by each patient j:

$$e_i = \sum_{j=1}^{n} P_{ijt} - C_{ijt},$$

where t indicates one of the particular types of treatment accepted, which is either high cost (H) or low cost (L). Specifically, a high-cost treatment is priced at P_H=115 points and costs a physician C_H=80 points. Low-cost treatment was priced at P_L=75 points and C_L=60 points. These parameters are chosen so that the physicians' profit for a high-cost treatment *(115-80=35 points*) is greater than that for a low-cost treatment (*75-60=15 points*). This is to ensure that physicians have incentives to overtreat patients,

[19] For the simplicity of the model and the experiment, the buyer must undergo the treatment with the second selected seller if he chooses to search for a second opinion. I assume that search costs increase dramatically with any further searches so that it would never be optimal for buyers to conduct further searches after receiving a second opinion.

which refers to the case in which physicians provide a high-cost treatment recommendation to patients with only a minor problem in this experiment. Moreover, in order to mimic real-world markets where overtreatment always leads to an efficiency loss, the above payoff parameters ensure that the physicians' gain from overtreatment is smaller than the net loss to the patients. Note that a physician's payoff is summed over the number of patients who choose to accept their recommendations. If any patient does not accept a physician's recommendation, the payoff in this period is zero. For simplicity, the researcher imposes verifiability in this experiment so that physicians can provide the announced treatment. All of the above information is common knowledge, except for the patient's types.

At the end of each period, both physicians and patients observe their own payoffs in the current period. In addition to the payoffs, physicians also observe the number of patients visited and the number of high-cost and low-cost treatments they performed.

3.2 Treatment Design

This study aims to explore the impact of reputation on physicians' overtreatment and examine whether the impact of reputation varies with the level of patients' search costs. To this end, a 2 × 2 factorial design is applied. The treatments differ in two dimensions: the possibility of reputation building and the search cost. I parameterize the search cost K to be equal to either 7 for the low search cost treatment or 14.5 for the high search cost treatment.

Based on the theoretical model in Mimra et al. (2016b), when reputation building is impossible, the low search cost condition (K =7) produces two types of equilibrium predictions: pure-strategy equilibrium and mixed-strategy equilibrium. In the pure-strategy equilibrium, physicians always overtreat patients with minor problems, and patients never search for a second opinion. In the mixed-strategy equilibrium, physicians' tendency to overtreat is balanced by the patient's tendency to search. Given the above parameters, one would observe that physicians overtreat patients with a minor problem in either 74.69% or 7.81%, and patients search for a second opinion in either 99.69% or 59.81%, respectively, when they receive a major treatment recommendation.

Under the high search cost condition (K = 14.5), if reputation is absent, the unique equilibrium overlaps with the pure-strategy equilibrium mentioned above, with complete overtreatment and no search. The intuition is that never searching for a second opinion (independent of physicians' overtreatment level) is optimal for patients when the search costs are prohibitively high. Thus, physicians' best response is to always overtreat patients with a minor problem.

The key treatment variation in this study is achieved by switching the reputation mechanisms on and off. For the no-reputation (NR) condition, physicians' ID numbers are shuffled after each market period. In these treatments, patients cannot distinguish between physicians; therefore, they can only choose randomly in the following market period. In the conditions with reputation (R), physicians' IDs are fixed between periods.[20] Starting from the second period, patients can also use their own past interactions with physicians in their first and second searches when making decisions[21]. *Table 1* summarizes the four treatments.

Table 1 Experimental Set-up: Conditions

	Low Search Cost (7)	High Search Cost (14.5)
No Reputation	NR_7	$NR_{14.5}$
Reputation	R_7	$R_{14.5}$

As mentioned in the introduction, all treatments in this study allowed patients to search for a second recommendation. This is because searching for a second recommendation is the only opportunity for buyers to assess a seller's dishonesty in a credence good setting. Instead of removing the opportunity to search, my design implements the high search cost condition, which theoretically prevents all patients from conducting searches but still leaves some flexibility, so it is possible to observe choices that do not conform to the theory.

3.3 Experimental protocol

A between-subjects design was implemented such that each subject participated in only one of the four treatment conditions. There were four sessions for each treatment, with a total of 128 subjects. Subjects were recruited from the University of Massachusetts Amherst subject pool using ORSEE (Greiner, 2015). The experiment was conducted using an online version of z-Tree (Fischbacher, 2007;

---

[20] Note that in all of the treatments, patients' IDs are shuffled at the beginning of each period. This set-up prevents physicians from engaging in complex and strategic play with certain identified patients.

[21] A rational buyer who accumulates a private history over the course of the game is always aware of his history. However, participants in the experiment might forget or misremember parts of their histories. I, therefore, display the private history in this step of the stage game as a reminder.

Duch et al., 2020).

At the beginning of each session, participants were invited to a zoom meeting room with their audio and video turned on. They were asked not to leave the meeting room until the end of the session. The experimenter read the instructions aloud, followed by a set of control questions to ensure that everyone understood the instructions (see Appendix B.1 for the instructions and B.2 for the control questions). During the experiment, participants were assigned to an individual breakout room so that the experimenter could assist them privately. After the experiment, a short questionnaire was used to collect demographic information (see Appendix B.3).

The payment includes a show-up bonus of $5 and the cumulative payoff from the decisions made in all 20 periods. The participants received their payoff with an Amazon eGift card[22] based on the exchange rate of 50 points = $1. The average payoff per participant was $16.47. The average session length was approximately one hour.

4. Hypotheses

In the analysis, four aspects are of prime interest: (i) whether the reputation mechanism reduces the level of physicians' overtreatment, (ii) whether the reputation mechanism reduces patients' search activities, (iii) whether introducing the reputation mechanism boosts market efficiency, and (iv) whether the impact of reputation depends on the level of search costs.

Based on the findings in the literature, I present the following hypotheses regarding how reputation impacts rates of overtreatment, patient search, and efficiency. I also hypothesize how the level of search costs affects the role of reputation.

Hypothesis 1: The Reputation Condition (Relative to no Reputation) will: a) decrease rates of overtreatment, b) decrease patient's need to search for a second opinion, c) improve market efficiency.

The availability of patient search provides an efficient way for patients to monitor physician honesty and makes the reputation mechanism more informative. In repeated interactions, physicians'

[22] The gift card was sent to each participant within three hours after the session ended.

decisions to overtreat not only induce their patients to search for second opinions and undergo treatment with their competitors but also risks future business with their patients once their dishonesty is detected. The expectations of future businesses further incentivize physicians to provide treatment recommendations truthfully. Again, in this game, a rational patient who receives a low-cost treatment recommendation should accept it with certainty because liability is applied to the market. Under reputation conditions, patients are more likely to receive a low-cost treatment recommendation on their first visit, and as a result, their need for a second opinion decreases. Finally, less overtreatment and fewer patient searches in the market lead to higher efficiency.

Hypothesis 2: The reputation effects will decrease with the increase of search costs.

Under high-search-cost conditions, patients are less willing to conduct a second search when they receive a high-cost treatment recommendation during their first visit. In this case, the chance of not getting caught and not being punished decreases the expected penalty, which increases physicians' dishonesty. With increased search costs, the reputation mechanism becomes less informative and effective in correcting physicians' incentives for overtreatment. An extremely high search cost will eventually remove patient search as well as reputational effects from the market, turning the game back to the conditions with reputation, as in Delluck et al. (2011) and Mimra et al. (2016a), where a buyer who goes to a seller must be treated by the seller.

5. Results

The following section examines the impact of the reputation mechanism on subjects' decisions in the credence goods market. The study conducts its analysis based on three main outcomes of the market: (1) physician overtreatment, (2) patient search, and (3) market efficiency. I first report treatment comparisons for each outcome variable using two-tailed Mann-Whitney U tests. For these tests, I consider decision dependency by taking the average of the measures across all individuals and all market periods as one independent observation. I accompany the nonparametric results with panel data analysis using random-effects regressions clustered at the market level. I also control for basic subject demographics and decision time trends in the regression results.

5.1 Overtreatment

I first turn to whether the reputation mechanism reduces the level of market overtreatment. Overtreatment is defined as a situation in which a minor problem receives a high-cost treatment recommendation.[23] The variable takes a value of 1 if overtreatment occurs and 0 otherwise. There are two distinct methods to measure the level of physicians' overtreatment. First, I analyze the "overtreatment strategy" based on all the recommendations submitted by a physician in her strategy profile. Recall that in the experimental design, a physician needs to make recommendations for each patient, even if the patient does not end up selecting her. This strategy design allows me to capture how many overtreatments were *intended* by physicians. This produces a clear indicator of a physician's honesty level. Second, I measure the "actual overtreatment" level experienced by the patients. It factors in the patients' choices of physicians. Under the above definitions, "actual overtreatment" occurs less frequently than "overtreatment in strategy," the difference of which reflects the ability of patients to detect dishonest recommendations and avoid overtreatment by selecting a different physician.

5.1.1 Overtreatment in Strategy

Figure 1 shows the average level of overtreatment in strategy across all four treatments and between the two reputational conditions (pooled across two search costs). Physicians choose to overtreat patients 52.54% of the time under conditions without reputation, a highly comparable result to that found in Mimra et al. (2016b, 52.7% overtreatment). When the reputation mechanism is introduced, physicians significantly reduce their tendency to recommend major recommendations for minor problems: only 36.76% of the recommendations are overtreated (R_7 & $R_{14.5}$ vs. NR_7 & $NR_{14.5}$, p=0.009). This provides strong support for Hypothesis 1.

Next, I examine the reputation effect under different search costs. The main result is largely driven by the significant difference in the low-search-cost condition (R_7 vs. NR_7, p=0.043). Under the high search cost condition, the effect of reputation on overtreatment is still negative but only statistically significant at the 10% level ($R_{14.5}$ vs. $NR_{14.5}$, p=0.083). This result confirms Hypothesis 2 on the interaction between reputation and search cost: reputation's effect is significantly reduced with a high search cost compared with a low search cost. As mentioned earlier, the intuition behind this is that

[23] The measure takes a null value if a buyer is assigned a major problem in a particular period.

when the search cost is prohibitively high, such as in the case of $R_{14.5}$ and $NR_{14.5}$, patients would be less willing to conduct costly searches. This, in turn, reduces the chance of detecting physician dishonesty and slows the reputation-building process. In Section 5.2, we examine the hypothesis regarding the effect of search costs on search frequency.

[Insert Figure 1 here]

The random effects regression results in *Table 3* confirm that physicians are less likely to overtreat patients when a reputation mechanism is present. In particular, under full specification (Model 3), the percentage of overtreatment drops by 10.7% when the reputation mechanism is introduced. The effect is statistically significant at the 1% level, with or without additional controls on basic demographics and the market period. These findings are consistent with hypothesis 1. However, the interaction effect between the two treatment dimensions is not statistically significant at the 10% level.

[Insert Table 3 here]

5.1.2 Actual Overtreatment

The level of actual overtreatment differs from the overtreatment decisions submitted by physicians in the strategy format, as the former are also factors in patients' active choices over physicians. Under conditions without a reputation, physicians overtreat patients 38.55% of the time. The results are also comparable to the findings of Mimra et al. (2016b, 36.46% overtreatment). As shown in *Figure 2,* patients are significantly less likely to be overtreated in conditions with a reputation than in those without. The average level of overtreatment drops to 21.67% after the reputation mechanism is introduced ($p=0.015$). When examining the effect under two different search costs separately, I conclude that the effect is mainly driven by the reduction in actual overtreatment in the low search cost condition (R_7 vs. NR_7, $p=0.017$). There is no significant difference in the actual overtreatment level between $R_{14.5}$ and $NR_{14.5}$ ($p=0.191$). Figure 2 reveals that the lack of statistical significance in the high-search-cost condition is driven by the high noise level in the decision data.

[Insert Figure 2 here]

The results from the random-effects probit regressions, as shown in Table 4, further confirm the significant and negative impact of reputation on the level of actual overtreatment. The marginal effect evaluated while holding all other variables at their mean suggests that patients are 21.6% less likely to receive overtreatment when reputation-building is possible. When comparing this result with the reputation effect on overtreatment in strategy, as shown in Section 5.1, I conclude that about half of the actual overtreatment is due to the change in recommendations selected by the physicians: They are 10.8% less dishonest on average. The other half of the effect results from patients being able to identify honest physicians and select their recommendations. Finally, I do not observe any interaction effect between the two treatment dimensions.

[Insert Table 4 here]

5.2 Patient search

I also use two ways to measure patients' search activities: conditional and unconditional search rates. Conditional search is only defined for cases where a patient receives a high-cost recommendation. If the patient chooses to search for a second opinion *conditional on* receiving a high-cost recommendation, the measure takes the value of 1, and 0 otherwise. Hence, the patient's conditional search rate is calculated as the count of the patient search divided by the total number of patients in all market periods *who received a high-cost treatment*. The conditional rate is a clean measure to capture patients' trust in a physician when faced with the ambiguous case of a high-cost recommendation[24]. The *unconditional* search rate is defined by dividing the total number of searches by the total number of patients in all market periods, regardless of the type of recommendation received in the first search. This measure reflects the amount of inefficient search cost incurred but does not take into consideration the number of high-cost treatments recommended by physicians. This is important because if a particular treatment entails more high-cost treatment recommendations, it is natural to expect more frequent searches and higher unconditional search rates. In summary, the conditional search rate is a

[24] Recall, in our experimental setting, if a patient observes a low-cost recommendation, she can infer that the physician must be offering the proper treatment honestly.

cleaner measure of patient trust, while the unconditional search rate is more directly linked to the market efficiency calculation.

5.2.1 Conditional Search Rate

Figure 3 shows the average search rate conditional on the high-cost recommendations. Upon receiving a major treatment recommendation on their first physician visit, patients search less often for a second opinion in conditions with a reputation than in those without a reputation (51% vs. 55.4%). However, this reduction is not statistically significant (P=0.40). There are also no significant differences for R_7 vs. NR_7 and $R_{14.5}$ vs. $NR_{14.5}$. The results indicate that the presence of a reputation mechanism does not alter patients' trust in physicians. As the results in Section 5.1 show a significant improvement in physicians' honesty, it is surprising that patients do not seem to incorporate physician behavior change into their best response.

[Insert Figure 3 here]

The panel probit regressions (see Table 5) report the results for the two dependent variables. In the first three models, the dependent variable is whether or not a patient searches for a second opinion when receiving a high-cost recommendation. The regressions report a similar null result on the effect of reputation, which is consistent with the results from the nonparametric analysis. In Model 1, the results show that patients are less willing to search for a second opinion when the search costs are high, which is not surprising given the Law of Demand. In Models 2 and 3, adding the treatment interaction term and other controls, I find that the interaction effect is significant at the 10% level: patients are 21.55% less likely to search for a second opinion if the physicians' reputation is observable and the search cost is high. The results suggest that patients may perceive reputation and search as substitutes: patients rely more on the physician reputation system if the price for additional search increases, and vice versa.

5.2.2 Unconditional Search Rate

Figure 4 shows that reputation is vital in reducing costly patient searches. The search rate pooled over both search cost treatments drops from 36.25% without reputation to 25.47% with reputation. The effect is highly significant (p=0.02). Clearly, the reduction is mainly driven by the decisions made in the high search cost condition (p=0.042). Under the low search cost condition, buyers' search rate entails higher noise, which implies that the comparison of search rate between reputational conditions (R_7 vs. NR_7) is no longer statistically significant. The above nonparametric results echo our hypothesis and the result observed on overtreatment in Section 5.1: high search cost leads to fewer searches, which makes reputational information difficult to accumulate.

[Insert Figure 4 here]

In Table 5, the dependent variable in the last three models is whether a patient searches for a second opinion. Accordingly, the results from Model 4 are consistent with those of the nonparametric test. Patients search for a second opinion 10% less often when they can observe their transaction histories. Moreover, they are less likely to search for a second opinion when search costs are high. However, the significance of reputation disappears when more factors are controlled for in Models 5 and 6. There is no interaction effect between reputation and search costs on patients' search activities.

An additional comparison between the conditional and unconditional search rates helps shed light on the source of the reputation effect on costly searches. By comparing Models 1 and 4, I conclude that the effect of reputation on reducing costly searches is caused by the desirable change in physicians' recommendations to more honest ones. This is not due to a change in patients' trust in physicians.

[Insert Table 5 here]

5.3 Market Efficiency

Market efficiency depends on actual overtreatment, search rates and the level of search cost. As reputation decreases overtreatment and reduces costly searches, I expect to see an improvement in market efficiency. To show this, I first define a relative efficiency measure by normalizing the sum of patients' and physicians' surpluses per market and per period in a [0,1]. Hence, a relative efficiency of 0 represents the minimal possible surplus of the market and corresponds to a situation in which physicians always overtreat and patients always search for a second recommendation; a relative efficiency of 1 represents the maximal possible surplus, which is achieved when physicians never overtreatment their patients and patients never search for a second opinion.

As shown in Figure 5, market efficiency is significantly higher in conditions with a reputation than in those without. Relative market efficiency rises from 56.09% to 73.52% on average when the reputation mechanism is available (p=0.002). The efficiency gain remains significantly positive after controlling for the level of search cost (p=0.021; p=0.021). The highest efficiency is observed for R_7 (75.08%). This is not surprising, as the low search cost and reputation reduce wasteful searches and overtreatment, respectively.

[Insert Figure 5 here]

The results from the regression models in Table 6 confirm the nonparametric findings and are robust to the inclusion of treatment interaction terms and demographic controls. The presence of reputation significantly improves relative market efficiency by 17.22%. In addition, the results of Model 1 reveal a weakly significant negative relationship between search costs and market efficiency, but the effect disappears after I add more controls.

[Insert Table 6 here]

6. Conclusion

Asymmetric information in healthcare services, a salient example of credence goods, often leads to inefficient market outcomes, as physicians exploit their informational advantage by recommending unnecessary high-cost treatments. This study implements a lab experiment to examine how reputation and patient search must be incorporated to improve the detection of dishonest physicians, build meaningful reputation records, and ultimately increase market efficiency.

Previous experiments on credence goods studied the effect of providing patients with past transactions but required patients to accept any treatment recommendations and forbid patient search. As a result, patients have little chance of verifying whether a past recommendation is an overtreatment or a proper one. To tap into the full potential of the reputation mechanism, this study relaxes this assumption and allows patients to conduct a second search by a different physician. Complementing Mimra et al. (2016b), the results demonstrate that physicians' reputation concerns motivate them to provide treatment recommendations more truthfully in repeated interactions. The reduction in physician overtreatment also greatly reduces patients' need to conduct costly second searches. Taken together, market efficiency is dramatically improved after reputation is introduced. Here, reputation is induced through repeated interactions and personal histories.

Moreover, this study explores how reputation effects vary with search costs. The hypothesis is that reputational incentives to mitigate overtreatment are weaker when patients search less frequently because of high search costs. The findings exhibit some supporting evidence for the differential reputation effect under different search costs from the nonparametric analysis. When the search cost is high, physicians' actions are less likely to be affected by the presence of a reputation mechanism. This result has important policy implications: patients with higher copays and deductibles for their health insurance suffer a higher search cost. Those who cannot afford to search for a second opinion may receive a disproportionate number of improper diagnoses and expensive treatment recommendations. This may further exacerbate the current public concern regarding healthcare inequality in society.

An interesting avenue for further research would be to analyze whether an alternative format for physician reputation, such as patient ratings and physician aggregated diagnoses, could result in additional improvements in physicians' honesty. Moreover, the current design assumes that both physicians and patients are homogenous in their incentives to provide recommendations and conduct searches. This is unlikely to be the case in realistic scenarios. In fact, previous theoretical studies have entertained the idea that healthcare service providers have different levels of competence and specialize in different treatments (Wolinsky, 1993). Mixing heterogeneous physicians and patients in the same market may generate an interesting extension of the current experiment.

Reference

Agarwal, R., Liu, C. W., & Prasad, K. (2019). Personal research, second opinions, and the diagnostic effort of experts. *Journal of Economic Behavior & Organization, 158,* 44-61.

Balafoutas, L., Beck, A., Kerschbamer, R., & Sutter, M. (2013). What drives taxi drivers? A field experiment on fraud in a market for credence goods. *Review of Economic Studies,* 80(3), 876-891.

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.

Bar-Isaac, H., & Tadelis, S. (2008). Seller reputation. F*oundations and Trends in Microeconomics*, 4(4), 273-351.

Brandts, J., & Potters, J. (2018). 17. Experimental industrial organization. *Handbook of Game Theory and Industrial Organization, Volume II: Applications*, 2, 453.

Brownlee, S. (2010). *Overtreated: Why too much medicine is making us sicker and poorer.* Bloomsbury Publishing USA.

Craig, D. (2005). *Rip off.* Original Book Company.

Currie, J., Lin, W., & Meng, J. (2014). Addressing antibiotic abuse in China: an experimental audit study. *Journal of development economics*, 110, 39-51.

Darby, M. R., & Karni, E. (1973). Free competition and the optimal amount of fraud. *The Journal of law and economics*, 16(1), 67-88.

Delattre, E., B. Dormont. 2003. Fixed Fees and Physician Induced Demand: A Panel Study on French Physicians. *Health Economics*. 12, 741-754

Dellarocas, C. (2006). Reputation mechanisms. *Handbook on economics and information systems*, 629-660.

Domenighetti, G., Casabianca, A., Gutzwiller, F., & Martinoli, S. (1993). Revisiting the most informed buyer of surgical services: the physician-patient. *International journal of technology assessment in health care*, 9(4), 505-513.

Duch, M. L., Grossmann, M. R., & Lauer, T. (2020). z-Tree unleashed: A novel client-integrating architecture for conducting z-Tree experiments over the Internet. *Journal of Behavioral and Experimental Finance*, 28, 100400.

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

Dulleck, U., Kerschbamer, R., & Sutter, M. (2011). The economics of credence goods: An experiment on the role of liability, verifiability, reputation, and competition. *American Economic Review*, 101(2), 526-55.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2), 171-178.

Fong, Y. F., Liu, T., & Meng, X. (2022). Trust building in credence goods markets. *American Economic Journal: Microeconomics*, *14*(1), 490-528.

Frankel, A., & Schwarz, M. (2014). Experts and their records. *Economic Inquiry*, *52*(1), 56-71.

Gerlach, H., & Li, J. (2022). Experts, trust and competition. *Journal of Economic Behavior & Organization*, *198*, 552-578.

Gottschalk, F., Mimra, W., & Waibel, C. (2020). Health services as credence goods: A field experiment. *The Economic Journa*l, 130(629), 1346-1383.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1), 114-125.

Gruber, J., M. Owings. 1996. Physician Financial Incentives and Cesarean Delivery. *The RAND Journal of Economics.* 27(1), 99-123.

Hu, W. (2017). Design Health Insurance for Overtreatment. *Available at SSRN 2970922*.

Huck, S., Lünser, G. K., & Tyran, J. R. (2012). Competition fosters trust. *Games and Economic Behavior*, 76(1), 195-209.

Iizuka, T. (2007). Experts' agency problems: evidence from the prescription drug market in Japan. *The Rand journal of economics*, 38(3), 844-862.

Johnson, E. M., & Rehavi, M. M. (2016). Physicians treating physicians: Information and incentives in childbirth. *American Economic Journal: Economic Policy*, *8*(1), 115-41.

Kerschbamer, R., & Sutter, M. (2017). The economics of credence goods–a survey of recent lab and field experiments. *CESifo Economic Studies*, 63(1), 1-23.

Lyu, H., Xu, T., Brotman, D., Mayer-Blackwell, B., Cooper, M., Daniel, M., ... & Makary, M. A. (2017). Overtreatment in the United States. *PLoS One*, 12(9), e0181970.

McGinnis, J. M., Stuckhardt, L., Saunders, R., & Smith, M. (Eds.). (2013). Best care at lower cost: the path to continuously learning health care in America.

Mimra, W., Rasch, A., & Waibel, C. (2016a). Price competition and reputation in credence goods markets: Experimental evidence. *Games and Economic Behavior*, 100, 337-352.

Mimra, W., Rasch, A., & Waibel, C. (2016b). Second opinions in markets for expert services: Experimental evidence. *Journal of Economic Behavior & Organization*, 131, 106-125.

Pieper, D., Hess, S., & Mathes, T. (2017). Compilation of second opinion programs in the German statutory health insurance. *Gesundheitswesen (Bundesverband der Arzte des Offentlichen Gesundheitsdienstes (Germany))*, 80(10), 859-863.

Rasch, A., & Waibel, C. (2018). What drives fraud in a credence goods market? – Evidence from a field study. *Oxford Bulletin of Economics and Statistics,* 80(3), 605-624.

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

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

Tadelis, S. (2016). Reputation and feedback systems in online platform markets. *Annual Review of Economics*, 8, 321-340.

Wagner, T. H., & Wagner, L. S. (1999). Who Gets Second Opinions? Once a technique for controlling inappropriate care, second medical opinions have become a controversial practice under managed care. *Health Affairs*, 18(5), 137-145.

Wolinsky, A. (1993). Competition in a market for informed experts' services. *The RAND Journal of Economics*, 380-398.

Appendix A: Tables & Figures

Table 1. Experimental Set-up: Conditions

	Low Search Cost (7)	High Search Cost (14.5)
No Reputation	NR_7	$NR_{14.5}$
Reputation	R_7	$R_{14.5}$

Table 2. Overview of results (means)

	Market with reputation		Market without reputation		P-values of MWU[1]		
Level of	R_7	$R_{14.5}$	NR_7	$NR_{14.5}$	**R** vs. **NR**[2]	R_7 vs NR_7	$R_{14.5}$ vs. $NR_{14.5}$
Overtreatment in Strategy	0.4060	0.3292	0.5143	0.5365	P= 0.0087	P=0.0433	P=0.0833
Actual Overtreatment	0.2083	0.2250	0.3792	0.3917	P=0.0147	P=0.0165	P=0.1913
Buyer Search	0.3188	0.1906	0.3969	0.3281	P=0.0203	P=0.1886	P=0.0421
Conditional Buyer Search[3]	0.6456	0.3791	0.5981	0.5147	P=0.4005	P=0.2454	P=0.1489
Market Efficiency[4]	0.7508	0.7196	0.6071	0.5146	P=0.0006	P=0.0209	P=0.0209

Notes: I analyze four independent markets under each experimental condition. Four sellers and four buyers interact in each market.

[1] Mann-Whitney U-tests for pairwise differences between conditions with matching groups of eight subjects over all 20 periods as one independent observation (note that besides the results displayed above, the level of both buyer search and conditional buyer search is significantly lower in R_7 than in $R_{14.5}$).

[2] Compare the conditions with reputation (R_7 and $R_{14.5}$) with those without reputation (NR_7 and $NR_{14.5}$).

[3] The level of condition buyer search measures the average rate of second opinions conditional on buyers receiving a major-treatment recommendation on their first visit.

[4] Market efficiency: calculated as $\frac{\text{Actual profit} - \text{Minimal possible profit}}{\text{Maximal possible profit} - \text{Minimal possible profit}}$

Table 3. Random effects panel OLS: the level of overtreatment in strategy

	Model 1	Model 2	Model 3
Reputation	-0.15781***	-0.10833***	-0.10683***
	(0.045)	(0.032)	(0.035)
High Search Cost	-0.02734	0.02214	0.02364
	(0.045)	(0.069)	(0.063)
Rep x High		-0.09896	-0.09857
		(0.086)	(0.083)
Period			-0.00105
			(0.003)
Male			0.04827
			(0.059)
Age			0.00229
			(0.007)
Take Economic Classes			-0.03868
			(0.057)
GPA			0.05006
			(0.047)
Constant	0.53906***	0.51432***	0.27741
	(0.031)	(0.028)	(0.205)
Observations	1,280	1,280	1280

Standard errors (reported in parentheses) are robust and clustered at the market level.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 4. Random effects panel Probit: the actual level of overtreatment

	Model 1	Model 2	Model 3
Reputation	-0.16673***	-0.16875***	-0.21546***
	(0.051)	(0.039)	(0.047)
High Search Cost	0.01330	0.01136	-0.01292
	(0.050)	(0.086)	(0.083)
Rep x High		0.00411	0.02604
		(0.096)	(0.094)
Period			-0.00305
			(0.004)
Male			0.02422
			(0.039)
Age			-0.01910*
			(0.012)
Take Economic Classes			0.03765
			(0.047)
GPA			0.04063
			(0.031)
Observations	960	960	960

Standard errors (reported in parentheses) were robust and calculated using the delta method.

The coefficients show the marginal effects of each variable, with all other variables held at their means.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5. Random effects panel Probit: patient search

	Conditional Patient Search			Unconditional Patient Search		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	(dy/dx)	(dy/dx)	(dy/dx)	(dy/dx)	(dy/dx)	(dy/dx)
Reputation	-0.03370	0.05981	0.06410	-0.10153***	-0.05870	-0.06830
	(0.063)	(0.074)	(0.082)	(0.036)	(0.049)	(0.055)
High Search Cost	-0.14847**	-0.05967	-0.02700	-0.09231***	-0.04952	-0.03829
	(0.063)	(0.103)	(0.114)	(0.035)	(0.055)	(0.064)
Rep x High		-0.18363*	-0.21553*		-0.08451	-0.09659
		(0.110)	(0.126)		(0.059)	(0.069)
Period			-0.00321			-0.00557**
			(0.003)			(0.003)
Male			0.00885			0.00142
			(0.071)			(0.042)
Age			-0.02463			-0.01784
			(0.019)			(0.012)
Take Econ Classes			-0.12612			-0.05899
			(0.092)			(0.059)
GPA			-0.04498			-0.00121
			(0.069)			(0.036)
Observations	724	724	724	1,280	1,280	1,280

Standard errors (reported in parentheses) were robust and calculated using the delta method.

The coefficients show the marginal effects of each variable, with all other variables held at their means.

*** p<0.01, ** p<0.05, * p<0.1

Table 6. Random effects panel OLS: relative market efficiency

	Model 1	Model 2	Model 3
Reputation	0.17436***	0.14371***	0.17220***
	(0.035)	(0.034)	(0.054)
High Search Cost	-0.06185*	-0.09251	-0.03437
	(0.035)	(0.060)	(0.059)
Rep * High		0.06132	0.01719
		(0.068)	(0.061)
Period			0.00480
			(0.004)
Male Average			0.10957
			(0.103)
Age Average			0.00872
			(0.015)
Take Economic Classes Average			-0.06629
			(0.088)
GPA Average			-0.19874
			(0.140)
Constant	0.59179***	0.60712***	1.03254**
	(0.030)	(0.033)	(0.465)
Observations	320	320	320

Standard errors (reported in parentheses) are robust and clustered at the market level.

*** p<0.01, ** p<0.05, * p<0.1

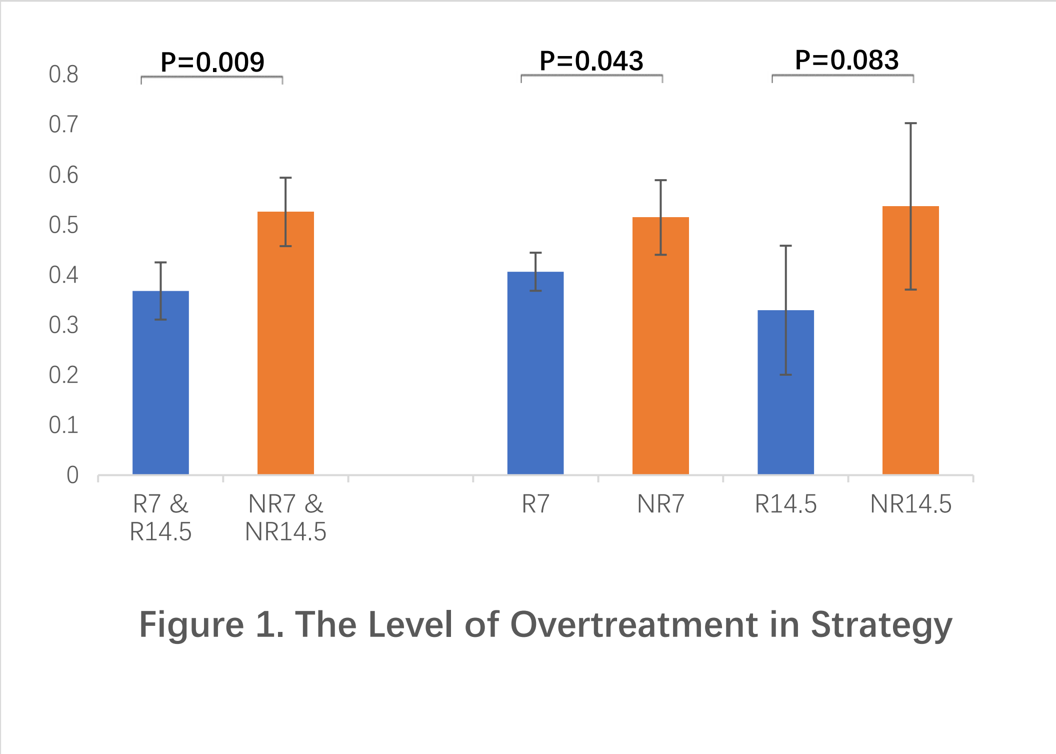


Figure 1. Level of Overtreatment in Strategy.

*p-value is based on Wilcoxon rank-sum test; error bars represent 90% confidence intervals.

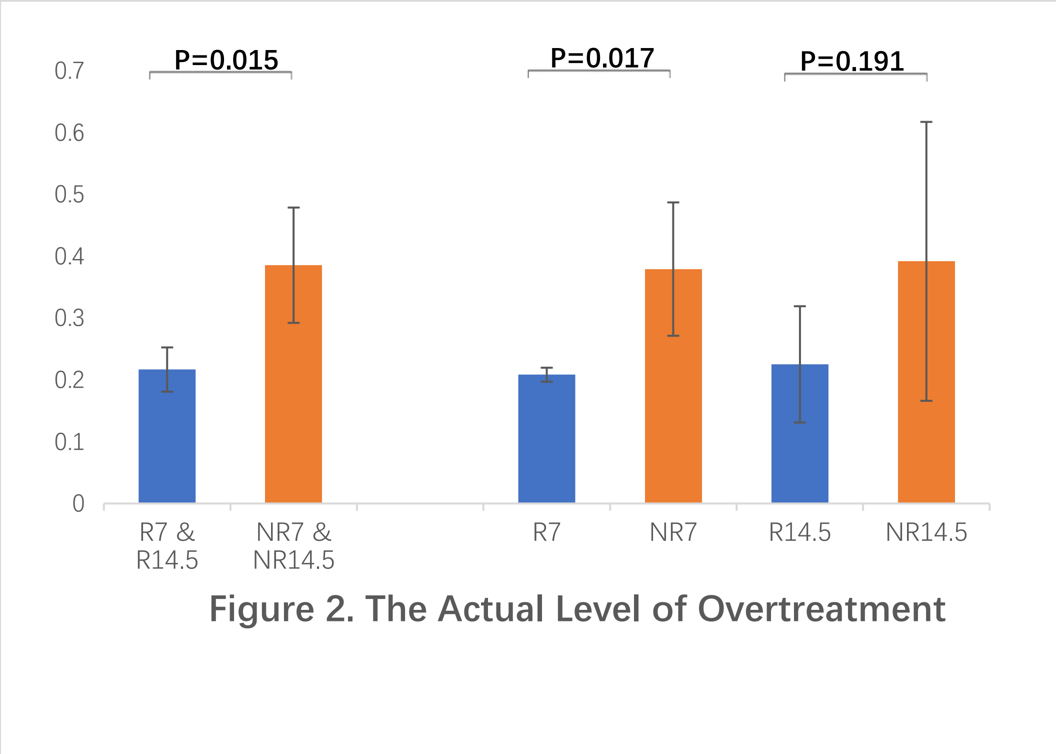


Figure 2. Actual Level of Overtreatment.

*p-value is based on Wilcoxon rank-sum test; error bars represent 90% confidence intervals.

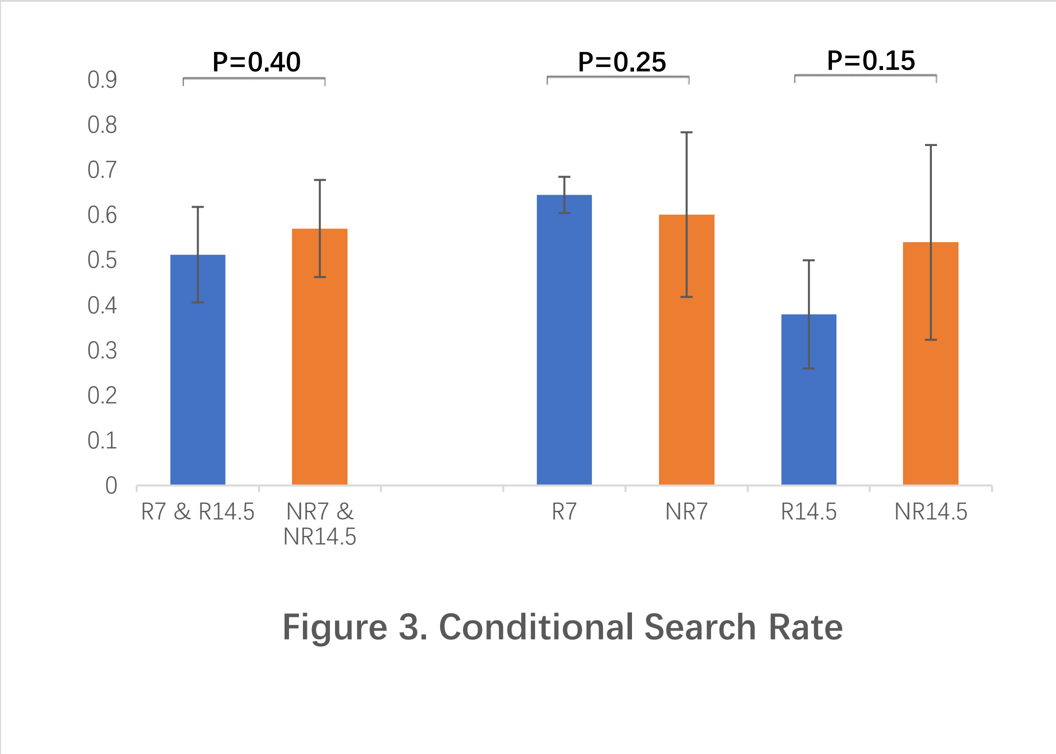


Figure 3. Conditional Search Rate.

*p-value is based on Wilcoxon rank-sum test; error bars represent 90% confidence intervals.

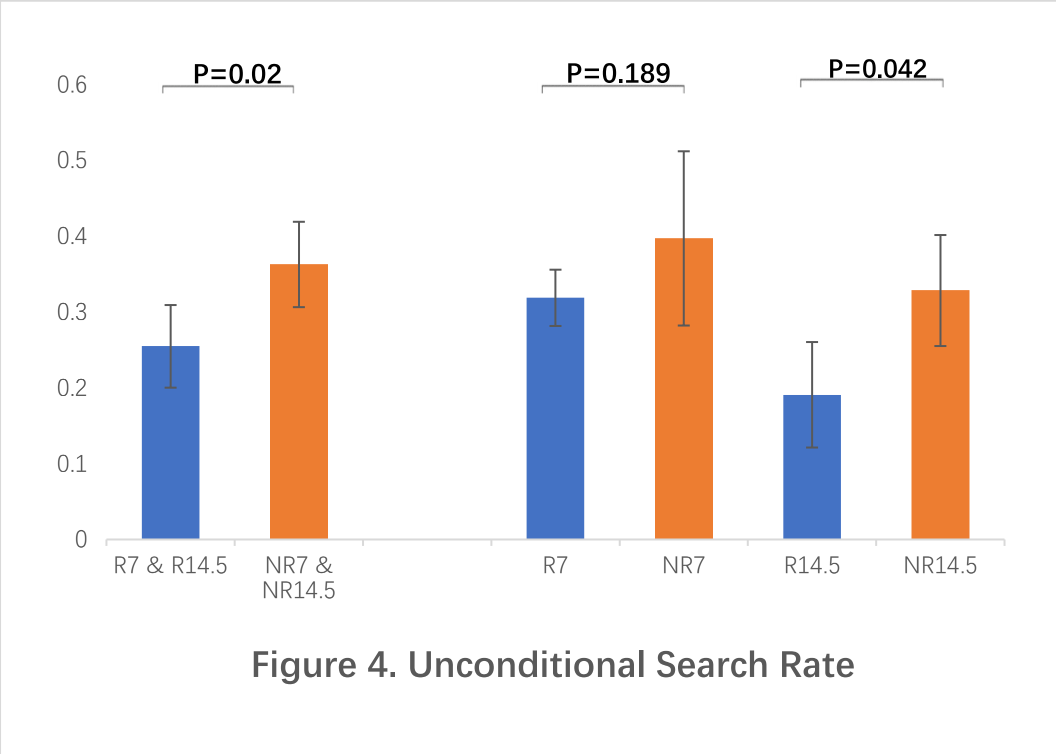


Figure 4. Unconditional Search Rate.

*p-value is based on Wilcoxon rank-sum test; error bars represent 90% confidence intervals.

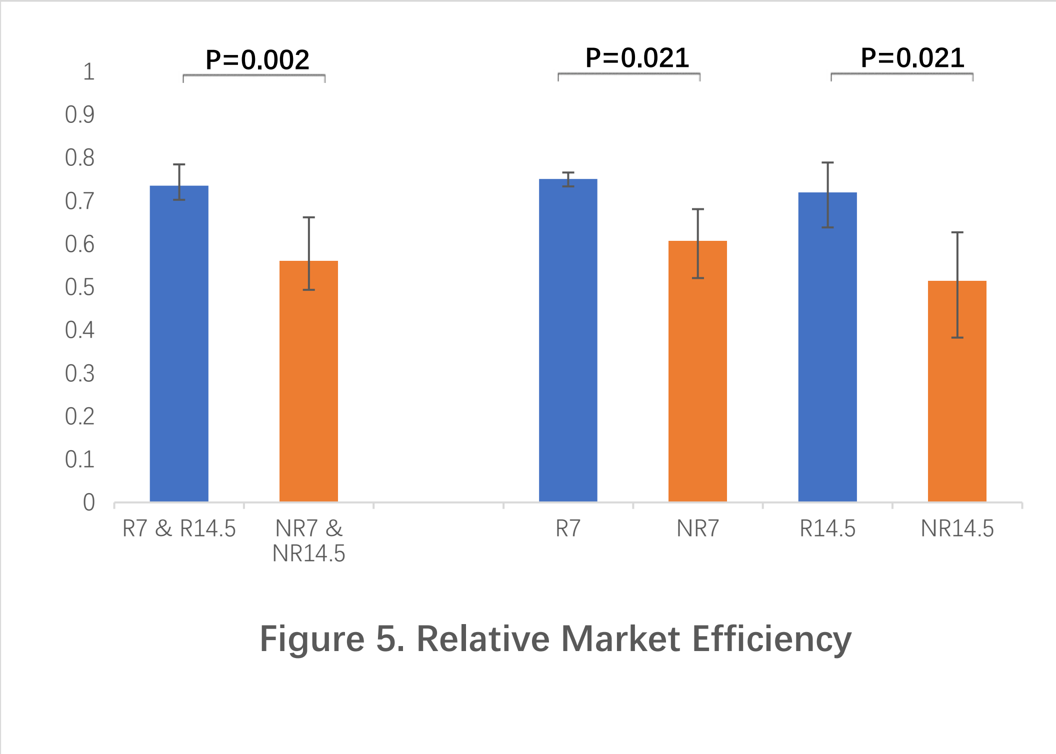


Figure 5. Relative Market Efficiency.

*p-value is based on Wilcoxon rank-sum test; error bars represent 90% confidence intervals.

Appendix B

B.1. Instructions for NR7 and R7[25]

To save space, I report the instructions for NR7, underline the parts that differ from those in R_7, and show the variation in brackets and underlined

Instructions

Thank you for participating in this experiment. In case you have a question, you can pop up your question through the chat. I will respond to your question after I finish reading the instructions.

You are paid $5 show-up bonus to be here on time. You can earn additional money depending on your decisions during the experiment. Below, the instructions specify how participants make decisions in today's experiment. To be able to earn more money, you will need to understand the instructions. After we finish reading the instructions together, you will be asked to take a short quiz to test your understanding. You WILL NOT be able to proceed to the experiment until you answer all quiz questions correctly. So please pay attention when I read the instructions.

The experiment is about the decisions of buyers and sellers in a market. Buyers need to choose a seller to purchase a service; sellers need to choose the type of service to offer to a buyer. In today's experiment, both sellers and buyers can earn "points" from their market transactions. At the end of the experiment, we will calculate your dollar earnings according to the following exchange rate:

50 points = $1

Market Roles

At the beginning of the experiment, you will be randomly assigned to be either a buyer or a seller. You will keep playing in the **SAME** role for the entire experiment. On the first screen of the experiment, you will see which role you are assigned to.

There will be 4 sellers and 4 buyers in a market. These eight people remain in the same market throughout the experiment. Once you are assigned to a market and given a role, you start interacting with other players in the market repeatedly. We call each of these repeated interactions a "period". There are 20 periods in total.

Within a period, you are given an ID number. Seller IDs can be either S1, S2, S3, or S4. Buyer IDs

[25] The only difference for NR_7 vs. $NR_{14.5}$ and R_7 vs. $R_{14.5}$ is that buyers need to pay 7.5 points more in R14.5 and NR14.5 when they search for a second opinion.

can be either B1, B2, B3, or B4.

Attention: Both Sellers' IDs and Buyer's IDs will be re-assigned randomly at the beginning of each period. That is to say, decision-makers behind each ID are different for each new period. However, the role each player plays is fixed throughout the experiment.

[In R7: Attention: Sellers' IDs are fixed throughout the experiment. The same ID number always represents the same seller. But: Buyers' IDs will be re-assigned randomly at the beginning of each period. That is to say, buyers behind each ID are different in each new period.]

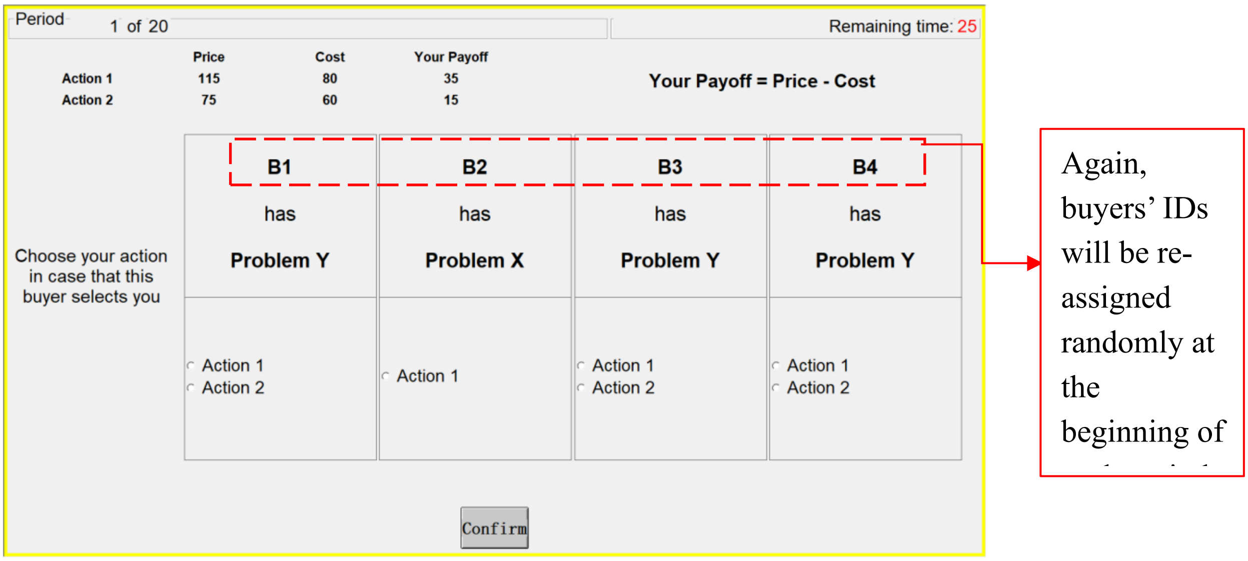
Decision Sequence

Once buyer and seller IDs are assigned, they make decisions in sequence to complete a market transaction. The instructions below detail the sequence.

At the beginning of a period, each buyer is randomly assigned a service problem: either Problem X or Problem Y. The type of problem is determined randomly and is unaffected by other buyers' problems. Problem X happens with a 25% chance. Problem Y happens with a 75% chance. Buyers do not know which type of service problem they have. To fix the unknown problem, the buyers have to interact with the sellers.

Unlike buyers, sellers can identify the type of problem each buyer has. Sellers can solve a buyer's problem by choosing one of the two possible actions, Action 1 or Action 2. Knowing that a buyer has Problem X, a seller MUST choose Action 1 to solve it. But if a buyer has Problem Y, a seller could choose to fix the problem using either Action 1 or Action 2.

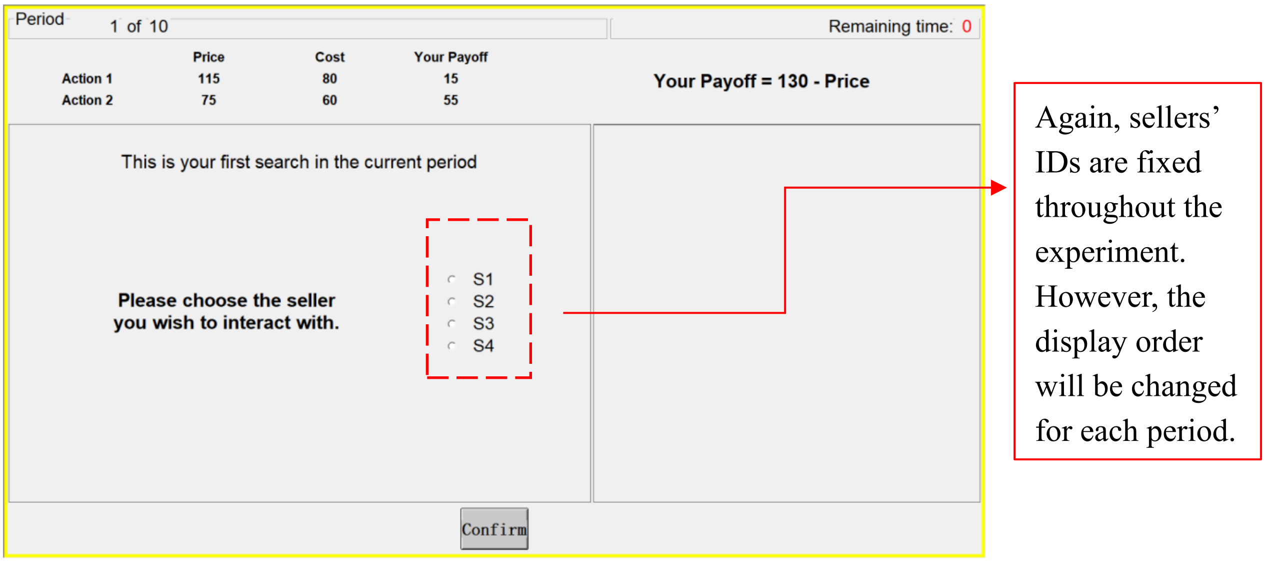
Moreover, the seller makes a total of four decisions, one towards each buyer. The following screenshot shows the decision interface for a particular seller. Note, since B2 has Problem X, this seller has no choice but choosing Action 1. For the other three buyers who have Problem Y, this seller could choose either Action 1 or Action 2.



(This interface is only viewed by sellers.)

Note that although a seller is choosing an action for each buyer, it does not necessarily mean that the seller will sell the service to all buyers. You can consider these seller's choices as "recommendations" to the buyers. Only in the case that a buyer chooses to interact with a particular seller and accept the proposed action, that seller's choice would be used to calculate his actual payoffs. For example, if a seller chooses Action 1 for all four buyers, but only B1 and B3 accept the actions provided by this particular seller, the seller's choice for B2 and B4 will not be used for payoff calculation.

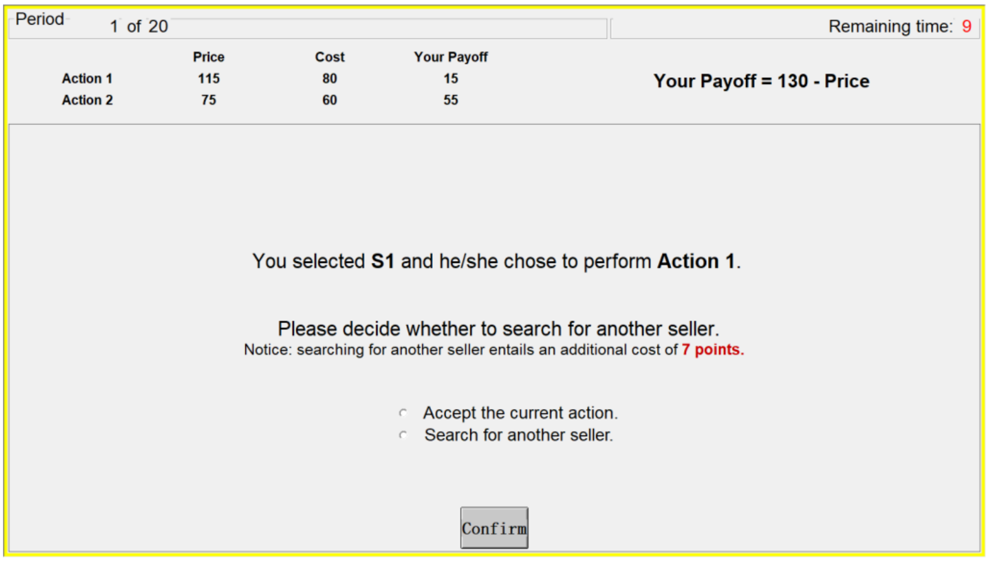
In this experiment, while sellers make their "recommendations", buyers need to select one of the four sellers in the market to interact with. Buyers make this selection without knowing what action each seller had chosen for him/her. The following screenshot shows the decision interface for a particular buyer.



(This interface is only viewed by buyers)

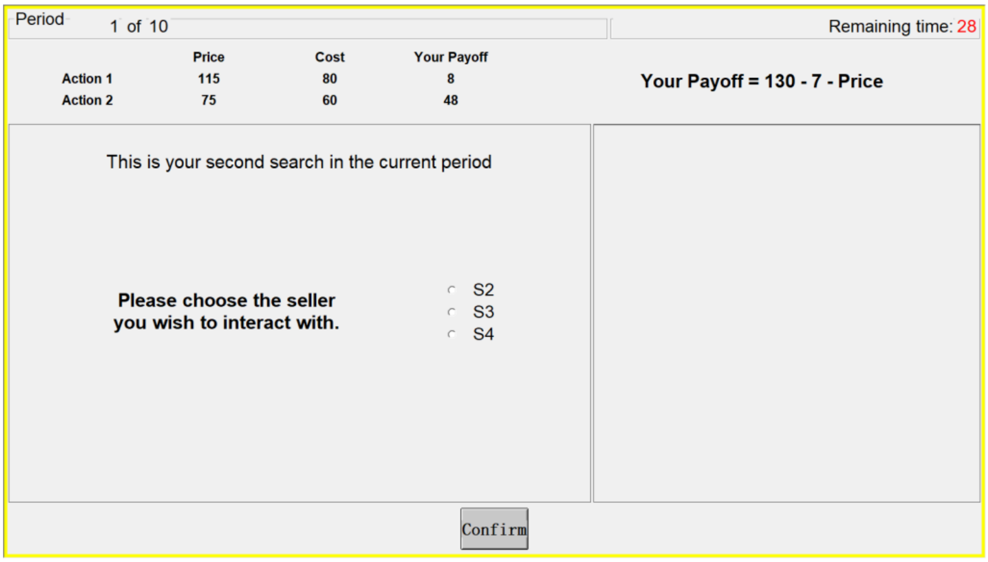
After the buyer selects a seller, s/he will observe the action offered by the selected seller. Now, the buyer has two options: (1) accept the action or (2) pay a cost of 7 points to search for another seller from the remaining three sellers. If the buyer chooses to accept the action, s/he pays the seller the price for the action and the period ends. If the buyer chooses to search for another seller, s/he pays a cost of 7 points and receives the action offered by the new seller. The action offered by this second seller is final. The buyer must accept it. The buyer does NOT have the option to select again, nor can s/he go back to accept the offer from the first seller.

The following screenshot shows an example of the decision interface for a particular buyer when s/he has selected a seller (in this case S1) and is asked to either accept the action (in this case, Action 1) or to search for another seller.



(This interface is only viewed by buyers.)

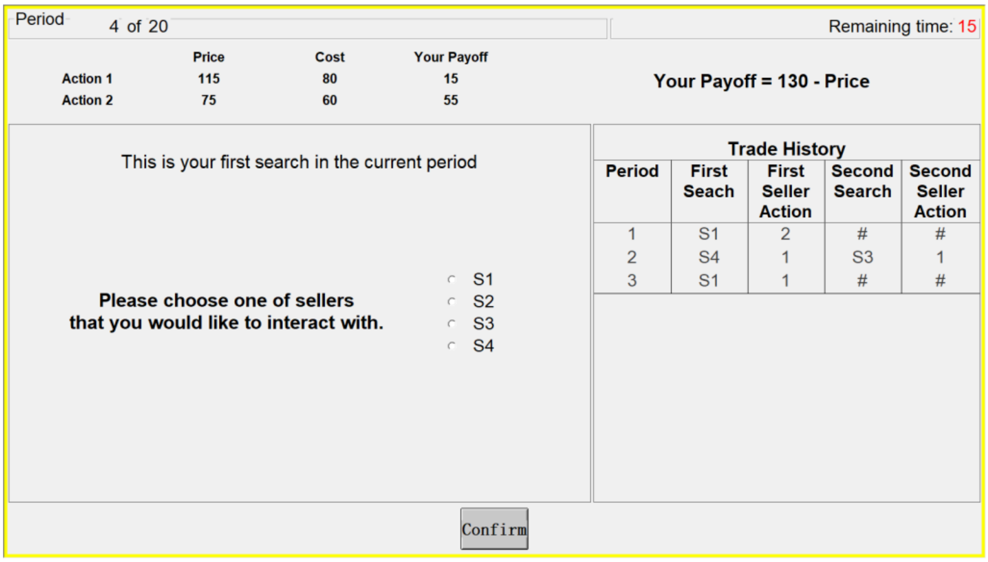
If the buyer selects "Search for another seller" in the screen shown above, s/he will see the interface below to select another seller from the remaining three sellers. Then the buyer receives the action offered by the new seller and pays the corresponding price. Notice that S1 is not on the option list. That is because this buyer chose not to accept the action offered by S1 in the previous decision screen.



(This interface is only viewed by buyers.)

[Only in R_7: Starting from the second period, each buyer can see a history table on the right half of the screen when s/he makes decisions. The table displays interactions from all previous periods, including information on the selected seller's ID and the action offered by that seller in both the first and the second search. If a buyer did not go through a second search, the related field will show a pound sign "#".

The following screenshot shows an example of the interface for a particular buyer during Period 4. The first row of the table shows you that this particular buyer knows that in Period 1, s/he selected S1 during the first search and S1 chose Action 2. This particular buyer accepted S1's action, therefore there is no information provided for the second search.]



[(This interface is only viewed by buyers in R_7.)]

In summary, in each period, sellers know the type of problem each buyer has and recommend an action to solve each problem. Simultaneously, each buyer chooses a seller without knowing what action that seller has chosen. The buyers then observe the chosen seller's action and decide on whether to accept the action or search for a second seller. Note: in case the buyer did not accept the first seller, that buyer must accept the action from the seller chosen in the second search. There is no more search after that. Also, the buyer cannot go back to the first selected seller once the second search begins.

Payoffs

You may wonder how Action 1 and Action 2 impact your payoffs. Here are the details.

For Action 1, a seller charges a buyer the price of 115 points and pays the cost of 80 points.

For Action 2, a seller charges a buyer the price of 75 points and pays the cost of 60 points.

Seller's earnings for a particular period are the sum of payoffs from all buyers who choose to accept that seller's action. The acceptance includes offers accepted during both buyers' first and second searches. For each accepted offer,

$$payoff_{seller} = price - cost.$$

If Action 1 is offered and accepted, the seller's payoff is 35 points (=115-80).

If Action 2 is offered and accepted, the seller's payoff is 15 points (=75-60).

If none of the seller's offers is accepted, the seller receives 0 points for the period.

A seller's total earnings are the sum of earnings across all 20 periods.

On the other hand, buyer payoff is calculated as follows:

$$payoff_{buyer} = \begin{cases} 130 - price_{first} & if\ the\ buyer\ accepts\ the\ first\ seller's\ action \\ 130 - price_{second} - 7 & if\ the\ buyer\ searches\ for\ a\ second\ seller \end{cases}$$

If Action 1 is offered and accepted in the first search, the buyer payoff is 15 points (=130-115).

If Action 2 is offered and accepted in the first search, the buyer payoff is 55 points (130-75).

If either of these actions is accepted in the second search, the buyer payoff will be 7 points lower. That is, 8 points for Action 1 and 48 points for Action 2.

A buyer's total earnings are the sum of payoffs across all 20 periods.

Notice that a buyer cannot find out what type of problem s/he has even after the payoff is determined and revealed (the buyer observes their own payoff, not the seller's payoff).

The information about the prices and costs of two different actions and your payoff will be listed at the top of the interface. These parameters are fixed throughout the experiment.

B.2. Control Questions for NR7 and R7

Please answer the following questions:

Q1: How many periods in total are there in this experiment?	Answer: 20
Consider the following hypothetical scenario: Suppose Buyer B1 selected Seller S2 as his first-choice seller and observed that S2 chose to perform Action 1. Then, B1 decided to switch to S3 in the second search and observed that S3 chose to perform Action 1 as well.	
Q2: How many points did S2 earn in this scenario?	Answer: 0
Q3: How many points did S3 earn in this scenario? Hint: The price of Action1 – The cost of Action 1	Answer: 35
Q4: How many points did B1 earn in this scenario? Hint: 130 - The price of Action 1 - 7 (search cost).	Answer: 8
Q5: Will B1 be able to go back to S2 and accept the offer from S2 if s/he chooses to search for a second seller? Yes/No	Answer: No

B.3. Questionnaires

1. Please enter your UMass e-mail address where we can send you the Amazon eGift card.

____

2. Which year were you born?

____

3. What gender are you identifying with?

- ◯ Male
- ◯ Female
- ◯ Other

4. Which academic cohort do you belong to as of this Fall semester?

- ◯ Freshman
- ◯ Sophomore
- ◯ Junior
- ◯ Senior
- ◯ Graduate Student
- ◯ Non-degree seeker

5. Which school are you majored in?

- ○ College of Education
- ○ College of Engineering
- ○ College of Humanities and Fine Arts
- ○ College of Information and Computer Sciences
- ○ College of Natural Sciences
- ○ College of Nursing
- ○ College of Social and Behavioral Sciences
- ○ Isenberg School of Management
- ○ School of Public Health and Health Sciences
- ○ Stockbridge School of Agriculture
- ○ Commonwealth Honors College
- ○ Undetermined

6. What is your current GPA?

- ○ 3.5 – 4.0
- ○ 3.0 – 3.49
- ○ 2.0 –2.99
- ○ Below 2.0

7. Have you ever taken any classes in economics?

- Yes
- No