Journal of Economic Behavior and Organization 183 (2021) 19–38

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

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

Recommendations in credence goods markets with horizontal product differentiation ☆

Katharina Momsen[1]

Department of Public Finance, Faculty of Economics and Statistics, University of Innsbruck, Universitaetsstrasse 15/4, Innsbruck 6020, Austria

ARTICLE INFO

Article history:
Received 5 May 2020
Revised 10 December 2020
Accepted 14 December 2020
Available online 7 January 2021

JEL classification:
D12
D18
D82

Keywords:
Credence goods
Horizontal product differentiation
Costly investment
Unobservable investment
Advice
Price competition

ABSTRACT

I experimentally analyze whether the introduction of watchdogs and the revelation of sellers' investment decisions can improve the market outcome in credence goods markets with horizontal product differentiation. Sellers can always give advice, yet they only observe consumers' valuations if they invest. I find that in the absence of watchdogs and with concealed investment decisions, both prices and investment rates are low and sellers give selfish advice. Each measure alone is not sufficient to improve recommendations; only the joint introduction has a significant positive impact on both quality and frequency of recommendations, but leaves welfare unaffected.

1. Introduction

When purchasing complex products, it may be hard for lay consumers to choose the variant which matches their preferences best. Take for example technologically advanced products like electronic devices: Although detailed information on the different components of smartphones, e.g. operating system, processor, screen and camera resolution, is available, many consumers cannot interpret these pieces of information and thus do not know which phone will be the best match for their needs. Having bought and used a certain phone, they will know how happy they are with their choice, but they will not be able to judge if another phone would have been an even better match for their preferences. The same reasoning holds for sports equipment such as running shoes: They differ, among other characteristics, in the materials used in and the construction of the sole. The optimal shoe should be selected depending on the runner's weight, running style, average running

☆ I am grateful to Henrik Orzen, Stefan Penczynski, Markus Ohndorf and two anonymous referees, seminar audiences in Karlsruhe (2016), Mannheim (2017) and Innsbruck (2017) as well as to the session participants of the IMEBESS (Florence, 2018), the ESA (Berlin, 2018) and the EEA (Cologne, 2018) for their valuable comments and suggestions.

E-mail address: katharina.momsen@uibk.ac.at

[1] The author was affiliated with the University of Mannheim (Germany) when the experiment was run. This research used generic funds provided by the Chair of Business Economics at the University of Mannheim. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

https://doi.org/10.1016/j.jebo.2020.12.018

distance and running surface. Before purchase, a lay runner cannot identify the shoe that would be the best match for their needs. Using the running shoe, they will experience how much they like it, but again, they do not know if the running experience would be better with a different shoe. Hence, for lay consumers many complex products qualify as credence goods (Darby and Karni, 1973). For more knowledgeable consumers, in contrast, the above-described products fall into the category of search or experience goods (Nelson, 1970). In the former case, consumers can identify their preferred version before purchase, whereas in the latter, they find out if they have purchased the right version by using the product.

Two separate definitions of credence goods go back to the seminal paper by Darby and Karni (1973). While, according to one definition, goods exhibit a credence characteristic if consumers do not know what they need, but observe the utility from what they get, the opposite applies when following the second definition. In this case, consumers know what they want or need, but they do not observe what they get nor do they experience the utility from what they get (Dulleck et al., 2011). For the products considered here, consumers always experience their utility from consumption, i.e. the products fall under the former definition. The difference between this type of credence good and experience goods lies in the information that is revealed through the experienced utility: Although consumers know how much utility they derive from such a credence good ex post, they still cannot judge if the purchased type or quality of the product is appropriate to their needs. In contrast, in the case of experience goods the ex ante unknown characteristics are revealed after purchase such that the experienced utility will inform consumers with certainty if they have purchased the needed version. There is hence a fine line between experience and credence goods as Darby and Karni (1973) point out.

So far, the existing literature (see Dulleck and Kerschbamer (2006); Kerschbamer and Sutter (2017); Balafoutas and Kerschbamer (2020) for overviews) has mainly investigated credence goods from a *vertical* product differentiation perspective with two versions of a good – a small version that satisfies consumers' needs with some probability and a large version which is always sufficient (Emons, 1997). Unlike suppliers, consumers do not know which variant they need, as it is the case for car repairs (Schneider, 2012) or taxi cab rides in unknown cities (Balafoutas et al., 2013). This information asymmetry can be exploited by sellers who can either provide insufficient treatment (undertreatment), provide a larger treatment than necessary (overtreatment) or charge for a larger treatment version than they have actually provided (overcharging).

In this article, I experimentally investigate credence goods markets with *horizontal* product differentiation. These markets deserve attention because many complex products qualify as horizontally differentiated credence goods to lay consumers and are thus not reflected by the existing literature. In these markets *mistreatment* instead of *over-* or *undertreatment* may reduce consumer welfare. Sellers can expect their mistreatment to remain undetected, because consumers will usually not find out if another version would have been better for them.

Let us return to the introductory examples: Of course, lay consumers could go through the troublesome process of informing themselves about all potential smartphone or running shoe options to figure out which version matches their needs best and thus transform a credence good into a search good, yet the search for information might be prohibitively costly. Hence, it might be more convenient for them to rely on the advice from retailers. The quality of the advice, however, may differ across retailers as some employ better-trained or more experienced salespeople who can better assess consumers' utility from each variant and can thus give better recommendations. As they require higher wages, their employment constitutes an investment for retailers. Observing the recommendation, it may be hard for consumers to judge if it stems from a qualified salesperson and whether the advice is genuine. After purchasing, the consumer will experience her consumption utility, but she may never find out if her choice was optimal because her valuations for the alternative options remain unknown to her. This lack of information may be exploited by sellers giving selfish recommendations. The aim of this article is to investigate how a market with cheap discounters giving unqualified, selfish advice can be transformed into an expert market where consumers receive reliable advice.

This article focuses on markets in which consumers and sellers interact repeatedly. Consider again the consumer electronics market: In many countries there are only a few different retailers which all offer the same variants of the same products. Each time consumers need a new electronic device – one day, it may be a smartphone, on another they are looking for a new computer –, they choose among the same retailers. The independent rounds in the experiment are meant to capture this feature of the market: The retailers' expertise is always specific to a round or, put differently, to the product category currently traded. When deciding which retailer to visit, consumers recall their experience from past purchases.

In the experiment, all sellers offer the same range of product variants with each seller facing different costs for each variant. Consumers differ in their valuations for the variants which are unknown to them. Sellers can help buyers to select a variant by giving them a recommendation. Without investing, sellers are unable to observe consumers' valuations for the different product versions; they only observe their own costs. Only if they invest, which is costly, can they also observe their clients' valuations. Investment is specific to the respective round, i.e. in each round sellers need to decide if they want to observe buyers' valuations. In the baseline treatment, consumers do not know which sellers have invested. As each subject is informed about her earnings, buyers can infer their valuation from the purchased variant, but they do not find out whether they would have had higher valuations for other variants such that they cannot judge their purchase nor the quality of the recommendation. I seek to identify factors which influence sellers' investment behavior and the quality of recommendations and thus determine whether sellers behave as discounters or as expert sellers. As one treatment variation I introduce watchdogs who test and rate the quality of sellers' advice. As another treatment variation I reveal sellers' investment decision to consumers.

The potential presence of watchdogs resembles occasional checks by consumer organizations which test products and services thoroughly to better inform consumers about product attributes and detect fraud towards consumers.[2] Sellers know that these organizations exist, but they do not know if and when their products or services will be subject to a test. It is only revealed ex post if they have been tested and how their product or service has performed. Similarly, in the experiment both buyers and sellers know that recommendations are subject to occasional checks, but they do not know ex ante when these checks take place. Introducing additional information to the market, watchdogs contribute to closing the information gap between buyers and sellers and enable sellers to build a reputation providing them with an incentive to improve their recommendations.

A visible investment decision constitutes an objective proof of sellers' qualifications such as certificates placed prominently in stores or offices. Outside the lab, requiring the public display of certificates or only allowing certified suppliers to enter the business can be easily-implementable forms of regulation. They inform consumers that the seller is capable of providing the appropriate service, yet they provide no information whether the seller actually makes use of his knowledge to increase consumer surplus. Thus far, labels and certificates have been investigated exclusively in vertically differentiated credence goods markets. Here, the literature has mainly come to the conclusion that the introduction of labels or information-providing activists decrease the credence goods problem and improve social (Feddersen and Gilligan, 2001; Roe and Sheldon, 2007) or consumer welfare (Bonroy and Constantatos, 2008), yet the findings presented here suggest that the existence of such improvements crucially depends on the form of the information-providing institutions as well as the type of information itself.

I find that in the absence of watchdogs and without observable investment decisions, prices are cheap, few sellers invest and most recommendations are selfish. In spite of this, consumers mostly follow the recommendations. The low price level ensures relatively high payoffs for consumers even though they usually do not purchase their most preferred option. The visibility of investment decisions encourages sellers to invest, but does not reduce the selfishness of the recommendations. Potential visits by watchdogs increase the consumer-friendliness of informed recommendations, but the share of investing sellers remains low. Only the joint presence of watchdogs and the revelation of investment decisions provokes frequent investment *and* consumer-friendly advice. In general, consumers are price-sensitive, yet in the treatment with watchdogs and observable investment decisions they accept higher prices. Buyers benefit from better recommendations in this treatment as their purchases better match their preferences. However, this positive effect on consumer surplus is offset by the rise in transactions prices which allows sellers to cover their investment costs. Hence, the introduced institutions transform a market with discounters into a market with better qualified sellers who give better recommendations and charge higher prices.

The remainder of this article is organized as follows: in Section 2, I will summarize the related literature. Then, I will explain the experimental design and in Section 4, I derive the behavioral predictions. I will present the results in Section 5. The last section concludes and the appendix provides a translation of the instructions as well as supplementary computations.

2. Related literature

There is a growing body of both theoretical and experimental literature on vertically differentiated credence goods (see e.g. Dulleck and Kerschbamer, 2006; Kerschbamer and Sutter, 2017; Balafoutas and Kerschbamer, 2020). The large-scale experiment by Dulleck et al. (2011) studies the determinants of efficiency in credence goods markets. Whereas in theory either liability or verifiability ensures efficiency, in their experiment only liability has a major impact. In horizontally differentiated credence goods markets, both liability and verifiability may have less bite since mistreatment is less easily detected than undertreatment and overcharging.

In a closely related theoretical article, Dulleck and Kerschbamer (2009) analyze competition between experts and discounters in credence goods markets. In contrast to discounters, experts have invested in diagnosis effort to identify the consumer's problem – a feature shared by this article. However, I do not allow for consumers visiting several sellers, such that they cannot free-ride on a – not necessarily honest – diagnosis of an expert and purchase from a cheaper discounter. Dulleck and Kerschbamer (2009) predict that, when switching costs are sufficiently low, experts randomize between always providing the cheap version and giving honest recommendations.

Dulleck et al. (2012) investigate the relationship between prices and quality. They find that good experts make use of high prices to signal their quality, but find no evidence of the opposite reasoning. Bad sellers set high prices to mimic good sellers, but deliver unsatisfactory quality.

In a theoretical paper, Pesendorfer and Wolinsky (2003) find that price competition reduces experts' incentive to exert effort and conclude that interventions to limit price competition may increase welfare. In an experiment, Mimra et al. (2016a) observe more fraudulent behavior under price competition than under fixed prices, arguing that the low price level under price competition inhibits quality competition. As the finding regarding the detrimental effects of competition is very robust, it is worth investigating which institutional settings can sustain a high quality of recommendations in the presence of price competition. While the literature has mainly focused on the option to gather second opinions

[2] Examples are the "Which?" Consumers' Association in the UK, Consumer Reports in the USA, Konsument in Austria and Stiftung Warentest in Germany.

(Pesendorfer and Wolinsky, 2003; Mimra et al., 2016b), I examine whether two information-providing institutions can improve the market outcome.

Both credence and experience goods exhibit strong information asymmetries between buyers and sellers, yet in the case of experience goods the information asymmetry vanishes after consumption, while it persists for credence goods. Inefficiencies related to experience goods result from undertreatment or no market interaction. Interpreting trust games as purchases of experience goods, Huck et al. (2012) find that reputation and competition increase trust. When trustees can determine the payoffs of rewarding or exploiting trust, trust and trustworthiness are lower compared to a setting with pre-determined payoffs (Huck et al., 2016). Hence, limiting price competition through regulation may also increase welfare in the context of experience goods.

This article also relates to the literature on ultimatum games which have been analyzed in the laboratory thoroughly (see e.g. Güth et al., 1982), but also several extensions have been taken to the lab. Mitzkewitz and Nagel (1993) study a version in which the size of the distributable payoff is only known to the proposer as it is the case with informed sellers in my experiment. Another version is the yes-no-game by Gehrig et al. (2007) in which the respondent does not observe the proposal, but in contrast to the dictator game she can still accept or reject the unseen offer. In Kriss et al. (2013)'s version of the ultimatum game, only the proposer can observe the allocation over which the parties bargain. They alter the possibilities of deception: With explicit deception, proposers communicate the allegedly drawn allocation together with their offer, whereas with implicit deception, proposers only state their offers. In my experiment, sellers can deceive buyers through their recommendations. A closely related article by Anbarcı et al. (2015) examines ultimatum games with messages and offers. In contrast to messages, which are cheap talk and can be false, offers state the actual split of the pie, but are only observed with some probability. Corazzini et al. (2014) introduce competition to such ultimatum games. In my experiment, the recommendations feature characteristics of the above-described messages.

3. Experimental design

Subjects interact on markets taking the roles of buyers and sellers. They are randomly assigned to their roles at the beginning of a session and keep their role throughout the 30 rounds. Under a fixed matching protocol, they are split into groups of eight, with a group representing a market of four sellers and four buyers. Whereas buyers cannot be identified by sellers, each seller is always represented by the same number. This feature allows sellers to build a reputation based on each buyer's private history.

In the existing experiments on credence goods, sellers usually offer two versions of a good or service: The large and possibly more expensive version always satisfies buyers' needs, whereas the cheaper version is only sufficient with a certain probability. In order to introduce horizontal product differentiation to a credence goods setting, in this experiment a larger number of product versions is available and sellers' marginal costs and buyers' valuations are uncorrelated. Each seller offers the same five variants, but variants change from round to round to ensure that learning corresponds to the situation in general. This design feature can be thought of as trading different goods in separate rounds. In each round, sellers know which variant maximizes their own profit, but they need to invest, which is costly, in order to observe which variant maximizes their client's utility. Buyers differ in their valuations for the different versions, which are unknown to them. Learning about the net utility of their purchase, but remaining uninformed about those of the other variants, they can only vaguely judge their purchase decision. Yet, they have the information necessary to compute the probability that the valuation for their purchased version constitutes the highest valuation in their sample.[3]

Each variant has different marginal costs which occur only when it is sold ('production on demand'). These marginal costs are integers randomly drawn without replacement from a discrete uniform distribution between 0 and 10 Experimental Currency Units (ECUs). As marginal costs are determined for each seller separately, they differ among sellers. In reality, these costs might depend on sellers' contracts with producers and differ depending on negotiated terms. Buyers have different valuations for the five product variants which are integers drawn without replacement from a discrete uniform distribution between 10 and 20 ECUs. Hence, the expected value of the highest valuation equals 19.[4] Note that both the interval from which sellers' costs and the interval from which buyers' valuations are drawn are common knowledge.

At the beginning of a round, sellers decide whether they want to invest in order to gain the ability to observe their customers' valuations. If a seller decides to invest, he incurs costs of 1.5 ECUs which need to be paid even if he cannot sell anything in the current round.[5] Then, each seller sets *one* price for all five product variants.[6] Although clearly unrealistic,

[3] See the appendix for more details on the computations.

[4] Based on the integral from which the valuations are drawn, one can derive the expected value of the largest integer in a draw: If five integers k were picked without replacement from an interval $\{1,\dots,n\}$ and I is the value of the highest integer, then $E[I]=\frac{k}{k+1}(n+1)$. For the valuations the interval $\{1,\dots,11\}$ needs to be shifted to the left by nine integers, as the lowest possible valuation is 10. Thus, $E[I]=\frac{5}{6}*12+9=19$, i.e. in expectation, buyers have a valuation of 19 for their most preferred version in each round.

[5] When sellers know buyers' valuations and include this knowledge in their recommendations, the *joint* welfare increases by 1.5 ECUs. With investment costs of 1.5 ECUs, efficiency preferences cannot affect behavior. This facilitates investigating if welfare is distributed differently depending on the sellers' investment decision.

[6] Sellers had to set their price within 35 seconds. If they did not set their price in time and ignored the reminder, the computer saved a price of 20 ECUs such that the experiment could continue. This mainly happened in the first five rounds which are excluded from the analysis. This feature was introduced

Table 1
Treatments.

		Observable investment decisions?	
		No	Yes
Watchdogs present?	No	Baseline	Observability
	Yes	Watchdogs	O + W

this design choice was made to keep the seller's pricing decision simple. Furthermore, it is supposed to resemble the average price *level* of each seller. As this article aims to investigate whether sellers become cheap discounters giving unqualified advice or expert sellers who give reliable advice, it is merely the price *level*, not specific prices which is of interest. When sellers make their pricing decision, the marginal costs of each product variant are depicted on their decision screens. Observing all prices in their market, buyers select a seller. As a reminder, they are shown a list with their past interaction partners and the respective payoffs from their purchases. Next, sellers are informed how many buyers they could attract. Sellers are not capacity-constrained. For each client, they need to decide which, if any, version to recommend. If they have invested, they can observe their client's net utility (valuation - price) for each product variant, and they are reminded of their respective margin (price - marginal costs).[7] Having received a recommendation, buyers can follow the advice, purchase a different variant or not purchase at all in this round. At the end of a round, all subjects are informed about their earnings. For a seller, the earnings are determined by their price from which the marginal costs of the sold variant are subtracted. If they found more than one customer, their earnings from the transactions are added. Investment costs are subtracted only once. A consumer earns her net utility (valuation - price) if she decided to purchase and zero if she has not interacted with a seller. In the next round, different products are traded, i.e. new costs and valuations are randomly determined and sellers need to decide if they want to invest to become on expert for the new product.

The experiment consists of four between-subject treatments implemented in a 2x2 design: I alter both the visibility of sellers' investment decisions and the presence of watchdogs who check the quality of sellers' recommendations (see Table 1). In the baseline treatment, no watchdogs are present and investment decisions remain unobservable. In the 'W' treatment watchdogs appear in randomly selected rounds to openly judge the quality of sellers' recommendations and in the 'O' treatment buyers observe the investment decisions before they select a seller. In the 'O+W' treatment, both watchdogs can occur and investment decisions are revealed.

In the treatments with watchdogs, both buyers and sellers know about occasional visits by watchdogs, but they cannot identify the relevant rounds *ex ante*. Note that it is not the products themselves, but the quality of the recommendations which is tested. I implement watchdogs by randomly selecting four test rounds from the 30 rounds of the experiment.[8] In a test round, one recommendation from each seller who was successful at attracting a positive number of buyers is selected randomly and graded.[9] The grading scheme is designed as follows: When proposing the variant yielding the lowest consumer surplus, the seller receives the worst possible grade (5). Analogously, when the seller recommends the variant which maximizes consumer surplus, his grade is a '1'. In case he has given a recommendation which neither minimized nor maximized consumer surplus, the grade is a '3'. A seller is not graded (0) if he has not found a consumer or has opted not to give a recommendation. Only the impact of the seller's recommendation on the consumer surplus is graded without taking his knowledge about the buyer's preferences into account. Sellers are informed about the presence of watchdogs and their grades after the recommendations have been judged. In the subsequent rounds, the grades of all four sellers in a market are depicted on the buyers' screens before they select their interaction partner. Note that grades from potential previous test rounds are also shown.

In the treatments with observable investment decisions, the investment decisions of all sellers in a market are revealed when buyers need to select their interaction partner, while they remain hidden in the other treatments. Sellers know that buyers observe their investment decisions both from the current round as well as from previous rounds.

Experimental procedure

The experiment was programmed in zTree (Fischbacher, 2007) and run at the University of Mannheim in the mLab in Spring 2017. In total 175 subjects participated who were recruited using ORSEE (Greiner, 2015). Most subjects were undergraduate students from all fields. At the beginning of a session, instructions were read out aloud to create common

as a precautious measure to ensure that the experiment could continue even if a computer stopped working or a subject refused to collaborate during a session.

[7] For simplicity, sellers represent a combination of both retailer and salesperson. In the role of the retailer, they make investment and pricing decisions while in the role of the salesperson they decide on recommendations.

[8] Subjects knew that test rounds were determined randomly, but they did not know the exact number of test rounds.

[9] One recommendation is selected randomly instead of grading each recommendation and awarding an average grade. In case of the latter an informed seller could guarantee a mediocre average grade by giving consumer-friendly recommendations if interests are aligned and selfish recommendations if not. Furthermore, testing and grading only one recommendation per seller adds to the realism of the design as watchdogs will only be able to check a subset of all recommendations given by a seller.

knowledge. Subjects had the possibility to ask questions, which were answered in private by the experimenters. As soon as all potential open questions had been answered, sessions started lasting on average 80 minutes. At the end of the experiment, ten of the 30 rounds were selected randomly for payment. The average earnings of these ten rounds were converted into Euros using different exchange rates for buyers and sellers to ensure comparable earnings. For sellers, one ECU was multiplied by five to get € 1, whereas for buyers the exchange rate was one. Additionally, buyers received a fixed fee of € 2.50 and sellers of € 5. On average, sellers earned € 12 and buyers € 14.40, such that the average payment was € 13.70. Subjects received their earnings privately and in cash.

4. Behavioral predictions

This article seeks to contribute to a better understanding of markets for credence goods with horizontal product differentiation. The virtual products qualify as credence goods – at least in most situations – as the consumer cannot judge ex post if she has acquired the product which best fits her preferences.[10] Only if her inferred utility equals 20 she knows that she has chosen the optimal product variant and the product becomes an experience good. If her inferred utility lies below 14, she is aware that she has not purchased her preferred version, but she does not know which version would have been a better match.

In order to derive behavioral predictions, I will first disregard the repeated nature of interactions and focus on the one-shot decision environment within a round starting with the baseline treatment. Here, I will proceed backwards along the decisions made by buyers and sellers. Unless explicitly stated otherwise, I assume that both buyers and sellers are rational payoff-maximizers.

Having received a recommendation, the consumer needs to decide whether to follow the advice. Since costs and valuations are uncorrelated, a selfish recommendation does not necessarily harm the consumer. Hence, *a priori* she is not made worse off by following the advice. If the seller has invested and gives consumer-friendly advice, she even benefits from purchasing the recommended good. The only explanation for not following the advice would be that she deliberately tries to lower the seller's payoff or believes that the seller recommends the version which minimizes her payoff. This behavioral pattern can only be explained by malevolence on the part of the buyer or the seller – which is highly unlikely in the current setting. Hence, following the advice is a weakly dominant strategy for the consumer.

Knowing that the consumer will follow his recommendation, the seller needs to decide whether to give a recommendation and which version to recommend. Considering a one-shot interaction, an uninformed rational seller should always give a recommendation and recommend the product with the lowest marginal costs. An informed seller seeking to maximize his profits should also give selfish advice. However, if he does not only care for his own payoff, but also for the buyer's, an altruistic seller might incorporate the buyer's payoff into his decision. In addition, reciprocity might play a role as the seller might seek to reward the buyer for choosing him over the other sellers.

Assuming sellers seek to maximize their own profits, consumers can expect recommendations to be selfish. Consequently, rational consumers will visit the cheapest seller. As all sellers offer identical product variants, Bertrand competition is likely to be fierce. Given that sellers expect consumers to follow their advice, the marginal costs of their least costly product variant matter for their pricing decision. Hence, prices will lie below the expected average marginal costs of 5 ECUs. Posting a higher price to signal that they have invested and will give consumer-friendly advice is not a credible strategy as uninformed sellers can mimic the signal. Sufficiently rational consumers will understand that a higher price does not mean that the recommendation will be superior and hence do not seek more expensive sellers.

Given the above reasoning, completely rational sellers will not invest in the baseline treatment.[11] They will charge low prices and give selfish advice which is nevertheless followed by consumers. Hence, the market will converge to a discounter market.

When the sellers' investment decisions are revealed and watchdogs are absent ('O' treatment), the market outcome will resemble the baseline treatment: Consumers will again follow the sellers' advice and sellers will give selfish recommendations since their incentive structure remains unchanged. As consumers can expect to receive selfish advice from both informed and uninformed sellers, they will prefer the cheapest seller. Given that potential investment costs are sunk when sellers make their pricing decision, both informed and uninformed sellers charge very low prices. Hence, sellers will not invest in the treatment with observable investment decisions. Investment can only be justified by behavioral reasons: If consumers seek informed, higher-priced sellers trusting that their advice will be better, profit-maximizing sellers should invest. Thus, the treatment with observable investment decisions will only differ from the baseline treatment if subjects deviate from the assumption of complete rationality.

In the 'W' treatment, the credence characteristic is transformed into an experience characteristic with some probability. Contributing to closing the information gap between buyers and sellers, watchdogs enable sellers to build a reputation. In a

[10] An inferred valuation of 19 is the highest valuation with a probability of 37.48% and 18 is the highest draw in 17% of the cases. A valuation of 14 as highest valuation is a highly unlikely event which occurs with a probability of 0.2%. See the Appendix for more details on the probabilities.

[11] It might also be the case that sellers refrain from investing because they want to remain ignorant about the impact of their recommendation on buyers' payoff in order to maintain a positive self-image while acting egoistically. This reasoning would be in line with the literature on moral wiggle room (see Dana et al. (2007)).

Table 2
Overview of average results.

	Baseline	O	W	O + W
Number of markets	6	6	5	5
Number of subjects	48	48	40	40
Posted price	6.27	6.34	6.64	6.56
Transaction price	4.51	4.98	4.74	5.61
Share of investing sellers	0.15	0.40	0.17	0.51
Share of purchase decisions with advice	0.96	0.93	0.86	0.94
Share of buyers following advice	0.52	0.51	0.46	0.74
Share of recommendations deserving best grade	0.21	0.24	0.25	0.35

one-shot scenario, however, this treatment variation does not influence behavior compared to the baseline treatment. Also the joint introduction of watchdogs and observable investment decisions remains without consequences.

If the repeated nature of the interactions is taken into account, reputation might play a role: Informed sellers might refrain from giving selfish advice in the situations when their selfish recommendation would induce a very low utility level for the consumer: If the consumer's utility lies below 14, she knows that the purchased version is not optimal and she might choose a different seller in the future. Hence, an informed profit-maximizing seller might try to avoid a bad reputation by giving advice which the consumer cannot detect as selfish. Similarly, he might give noticeably consumer-friendly recommendations sacrificing payoffs in the current period in order to build a good reputation.[12] The credence characteristic of the considered goods, however, severely impedes the possibility to build a good reputation and thus behavior in the baseline treatment as well as in the 'O' treatment will resemble the above-described one-shot decisions.

If watchdogs are present ('W' treatment and 'O+W' treatment), uninformed sellers might refrain from giving advice fearing that bad grades will deter future consumers from purchasing their products. Here, informed sellers might give consumer-friendly recommendations trying to build a good reputation. Building a good reputation, however, only makes sense for sellers if buyers take their grades into account when selecting a seller. Otherwise prices remain the only selection criterion. Note that it is a finitely repeated game: In the last round, it is always optimal for sellers to give selfish recommendations. Via backward induction, behavior in the finitely repeated game should be identical to the one-shot game: Sellers give selfish recommendations and do not invest. However, as subjects interact for 30 rounds, building a reputation might still pay off (see Huck et al., 2012). Overall, behavior in the treatments with watchdogs can only differ from the other treatments if the repeated nature of the interactions is taken into account.

Assuming rational decision makers, the behavior in both treatments with watchdogs should be identical as revealing the investment decision does not change sellers' incentive structure. Hence, potential differences in behavior must be attributed to behavioral patterns in decision making, e.g. demonstrably informed recommendations harming the buyer may have a more detrimental effect on the seller's reputation than uninformed selfish recommendations (see Conrads and Irlenbusch, 2013). Moreover, buyers might misinterpret the investment decision and sellers might anticipate this misinterpretation.

Note that, in the experimental context, a market with informed sellers who incorporate consumer valuation in their advice and a market with uninformed sellers are identical in terms of efficiency. Hence, the examined institutional measures would only affect the distribution of welfare.

5. Results

Table 2 provides a rough overview of the main decisions for each treatment.[13] Average posted prices are low in all treatments, yet they exceed 5 ECUs – the price at which a seller who has not invested can expect to break even if consumers choose randomly. As can be seen from transaction prices, consumers are very price sensitive but less so in the 'O+W' treatment. In the treatments without observable investment decisions, sellers are reluctant to invest. Buyers usually receive a recommendation in all treatments. About half of the recommendations are implemented in all treatments but the 'O+W' treatment where almost three fourth of the buyers follow the advice. The share of recommendations deserving the best grade is maximal in the 'O+W' treatment, but rather low on average.

[12] Theoretically, consumers can compute the probability that the purchased version constitutes the optimal fit to their preferences; it is, however, cognitively very demanding.

[13] Unless explicitly stated otherwise, I focus my analysis on rounds 6 to 30. Excluding the first rounds, I ensure that treatment effects are not polluted by potentially erratic choices at the beginning of the experiment. In 2% of the decision situations, at least one seller in a market failed to set a price. I also exclude these observations.

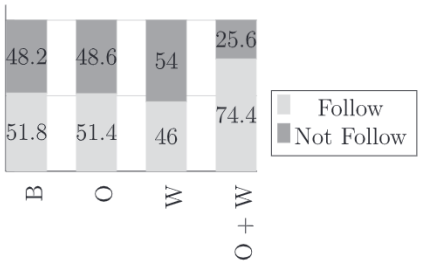


Fig. 1. Percentage of buyers following the recommendation.

Table 3
Regressions: propensity to follow the seller's advice.

	(1) No observability	(2) Observability
Investment		0.285***
		(0.069)
Investment * Watchdogs		−0.033
		(0.109)
Watchdogs	−0.064	0.197**
	(0.092)	(0.098)
Price	0.036**	0.034***
	(0.017)	(0.013)
Period	−0.002	0.006
	(0.015)	(0.012)
Period^2	−0.000	−0.000
	(0.000)	(0.000)
Constant	0.390**	0.167
	(0.178)	(0.149)
R^2	0.005	0.196
N	1057	1055

Notes: Dependent variable is a dummy indicating whether the buyer follows the advice. Output from random effects panel regressions. Cluster-robust standard errors in parentheses (on subject-level). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Results remain unchanged when clustering on matching-group-level.

5.1. Recommendations

Do buyers follow the recommendations?

Starting with the last decision in a round and proceeding backwards, I first analyze whether buyers follow the recommendations. As can be seen from Fig. 1 buyers follow the advice in about 50% of the situations in all treatments except for the 'O+W' treatment. In this treatment, 74% of the recommendations are implemented which constitutes a significant difference ($p = 0.015$) to all other treatments.[14] If buyers randomized over all five options, they would follow the advice in 20% of the situations. Hence, buyers take recommendations into account, with the extent to which they do so differing between treatments. According to the behavioral predictions, buyers should favor following the advice if they believe that they face an informed and honest recommendation; in all other cases, they should be indifferent between all five options – a reasoning which buyers seem to have (at least partially) understood.

In addition to these rational considerations, behavioral reasons related to the posted price may drive the decision to follow the advice: On the one hand, buyers might be more likely to follow the advice of a cheap seller seeking to reciprocate his action. On the other hand, they might interpret a higher price as a signal for better recommendations. To investigate which reasoning applies, I regress the propensity to follow the recommendation on the posted price and a dummy indicating whether watchdogs can occur while controlling for the round number (see Table 3). I only include a dummy for the seller's investment decision in the regression for the treatments with observable investment decisions (Regression 2). I further interact this dummy with the variable indicating the presence of watchdogs. I cluster standard errors – which are presented below the coefficients – on subject level. In order to keep the interpretation of coefficients simple, I present the outcome of a random effects panel regression. I find the investment decision of the interacting seller, the presence of watchdogs and the price to have a positive and significant effect in the treatment with observable investment decisions. The presence of watchdogs, however, does not have an additional impact on buyers' propensity to follow the advice of an informed seller. In the treatments with hidden investment decisions, only prices have a positive and significant impact on buyers' willingness to

[14] Unless explicitly stated otherwise, reported p-values are based on Wilcoxon rank-sum tests taking the average over all rounds and all subjects for each market as independent unit of observation.

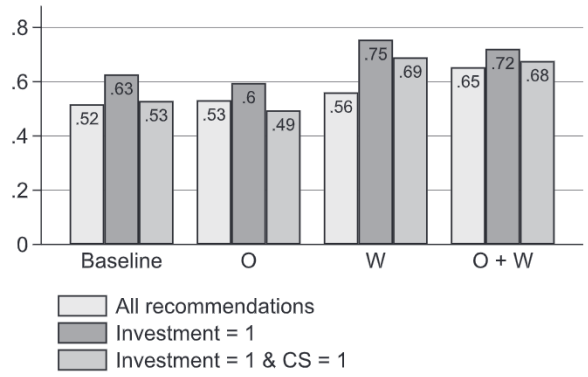


Fig. 2. Share of consumer surplus achievable by following the recommendation.

follow the advice. Hence, there is some evidence for buyers interpreting higher prices as signal for "better" recommendations while reciprocity considerations do not seem to play a role. Here, observed behavior contradicts the predictions based on the assumption of rationality on both market sides: Buyers should be aware that uninformed sellers can easily mimic the signal of a higher price. Furthermore, they should realize that informed rational sellers still have an incentive to give selfish advice.

Do buyers benefit from following the advice?

To analyze the impact of following sellers' advice on consumer surplus, I construct a variable indicating which share of the largest possible consumer surplus could be achieved by following the recommendation.[15] This variable is computed by subtracting the lowest possible consumer surplus from the surplus of the recommended option and dividing the result by the difference between the highest and the lowest valuation in the set of possible valuations in a round:

$$\text{Share of achievable consumer surplus} = \frac{\text{valuation}_{\text{recommendation}} - \text{valuation}_{\text{min}}}{\text{valuation}_{\text{max}} - \text{valuation}_{\text{min}}}$$

The lower the share, the less optimal is the recommendation. Fig. 2 depicts the average shares for each treatment, differentiating between all recommendations (first bar), recommendations given by informed sellers (second bar) and recommendations of informed sellers in situations with conflicting interests (conflict situations *CS*) between buyer and seller (third bar), i.e. situations in which the option maximizing seller's profits does not maximize consumer surplus. In the treatments without watchdogs, about half of the possible consumer surplus can be realized by following the advice. As valuations and costs are uncorrelated, this is the expected share reachable by following selfish recommendations. When watchdogs are introduced, however, following the recommendation can lead to a larger share of the possible consumer surplus. Considering only informed recommendations – in general and in conflict situations – this effect is more pronounced, but note that in half of the treatments consumers cannot identify informed sellers. While the revelation of the investment decision does not influence the share of consumer surplus achievable by following the recommendation, watchdogs have a significant positive impact. In light of the repeated nature of the interaction, the fact that watchdogs transform credence goods into potential experience goods induces informed sellers to invest in building a reputation by giving better recommendations.

When do sellers give recommendations?

In the vast majority of purchase decisions, buyers receive a recommendation (92.57%). Fig. 3 illustrates the four possible constellations of advice and investment for all purchase decisions in each treatment. Relative to the investment decisions, the segments of the bars depict the share of purchase situations in which buyers received and did not receive a recommendation. In the treatments with hidden investment decisions the majority of recommendations stems from uninformed sellers. The large share of informed recommendations in the 'O' treatment cannot be reconciled with rational actors on both market sides: Since incentives are unchanged compared to the baseline treatment, rational sellers should not invest. In the 'W' treatment, both informed and uninformed sellers are more likely to refrain from giving advice than in the other treatments potentially fearing the negative reputational consequences of bad advice. However, this reasoning was not very wide-spread.

What do sellers recommend?

Focusing on informed sellers, I now investigate which option they recommend. Fig. 4 illustrates which share of recommendations in a treatment is given by informed sellers (left) and which share is given by uninformed sellers (right).[16]

[15] As the decision to follow the recommendation is made after selecting a seller, this analysis abstracts from prices and only considers valuations.

[16] For each treatment, the two bars sum up to 100%, i.e. to all recommendations in the respective treatment.

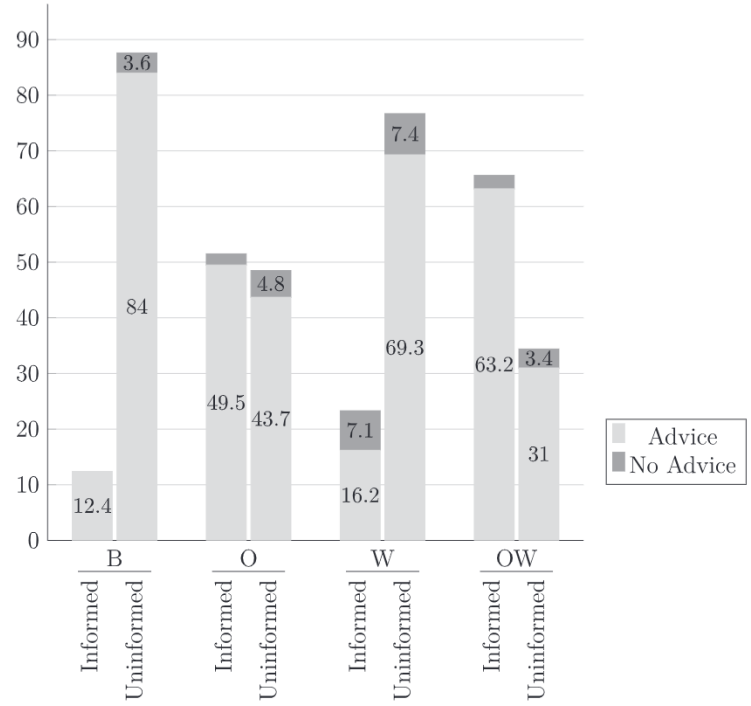


Fig. 3. Advice relative to the seller's investment decision.

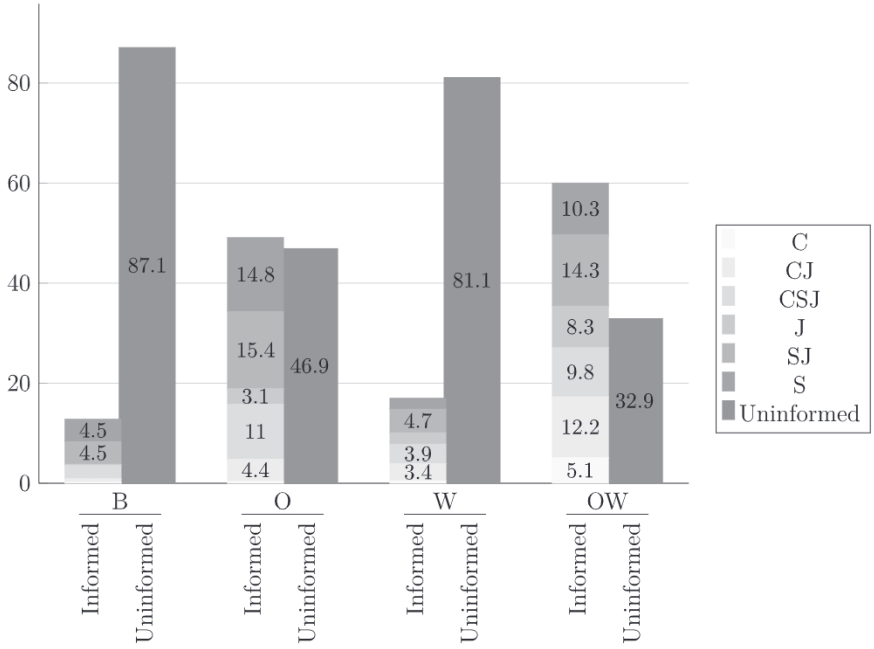


Fig. 4. Content of the recommendations.

Furthermore, it depicts what percentage of informed recommendations stands in line with the most obvious objectives: own profit maximization (S), maximization of consumer surplus (C) and maximization of joint welfare (J). Note that several aims may also be fulfilled simultaneously (CSJ, SJ, CJ). The segments of the respective left bars show what share of recommendations falls into each category. The right bar is not split up into different segments because uninformed sellers could not assess the impact of their recommendation on their client's surplus.

In the baseline treatment only about 13% of the recommendations stem from informed sellers and of these recommendations almost all maximize seller's profit (CSJ, SJ, S). As valuations and costs are uncorrelated, these recommendations happen

to maximize consumer surplus in roughly 20% of the situations (CSJ). There are only five out of 71 recommendations which maximize consumer surplus and are not optimal for sellers (C, CJ). About 63% of the recommendations are welfare-optimal (CSJ, SJ) – a share which is not significantly different from the expected share of welfare-optimal recommendations if sellers only maximized their own payoff.

When investment decisions are observable, about half of the recommendations come from informed sellers, but in only 30% of the cases sellers use their knowledge to maximize buyers' welfare (CSJ, CJ, C). There are very few recommendations which maximize only consumer surplus, but do not take into account overall welfare. Hence, most sellers seem to use their investment decision as a signal to attract clients. 78% of the informed recommendations aim at maximizing sellers' profits (CSJ, SJ, S) and almost 64% maximize the sum of buyer and seller surplus (CJ, CSJ, J, SJ). Although significantly more recommendations stem from informed sellers compared to the baseline treatment ($p < 0.01$), the relative frequency of the three considered objectives is practically unaffected – the former result contradicting and the latter result in line with the prediction of no difference between the baseline and the 'O' treatment.

When watchdogs are introduced, 18.92% of the recommendations are given by informed sellers. This does not constitute a significant difference to the baseline treatment. However, the pattern according to which sellers determine their recommendations appears to have changed: In the presence of watchdogs, 41.6% of the informed recommendations maximize buyers' surplus (C, CJ, CSJ) and only 57.1% are selfish (S, SJ, CSJ). Both the decrease in selfishness and the increase in the share of welfare-maximizing messages are significant at the 10%-level, while the change in the share of consumer-friendly recommendations is insignificant. As watchdogs enable sellers to build a reputation, it is not surprising that the quality of informed recommendations has improved. Instead, it is the large share of uninformed recommendations which cannot be reconciled with the predictions.

In the 'O+W' treatment, almost 70% of the recommendations come from informed sellers. With only 51% selfish recommendations, this treatment is the one in which informed sellers are most willing to give up their own profit in order to maximize buyers' surplus. 77% of the recommendations are welfare-optimal. The share of consumer-friendly recommendations lies slightly below the treatment with watchdogs only but above the treatments without watchdogs at 40%. Overall, observable investment decisions induce sellers to invest more often and watchdogs motivate them to behave less egoistically and also take buyers' surplus into account. These two effects taken together explain the results in 'O+W' treatment.

What do sellers recommend when their interest and their client's interest are not aligned?

In situations with a trade-off between seller and buyer surplus, most sellers do not refrain from giving advice.[17] To analyze seller behavior in conflict situations, I construct a variable indicating the intensity of the conflict of interests. I therefore subtract the lowest costs in the seller's set of cost realizations from the costs of the variant maximizing consumer surplus. The resulting difference in costs is divided by the difference between the highest and lowest costs.

$$\text{Conflict intensity} = \frac{\text{Costs}_{\text{CS max}} - \text{Costs}_{\text{min}}}{\text{Costs}_{\text{max}} - \text{Costs}_{\text{min}}}$$

I utilize this continuous measure of conflict intensity as explanatory variable in a regression of the share of consumer surplus reachable by following the recommendation (see Table 4). Note that a larger value of the variable indicates a stronger conflict of interests.

Conflict intensity has a significant decreasing effect on the share of consumer surplus reachable by following the recommendation: The stronger the conflict in interests, the less likely are sellers to recommend options maximizing consumer surplus. The presence of watchdogs has a positive, but insignificant impact. Interacting conflict intensity with the dummy variable indicating the presence of watchdogs, we observe a positive and significant coefficient. Hence in the presence of watchdogs, the decreasing effect of the conflict intensity is (partially) mitigated. The effect of observable investment decisions remains insignificant. This finding is in line with the behavioral predictions: The observability of the investment decision does not affect sellers' incentive to give good recommendations.

5.2. Purchase decisions

Buyers can base their purchase decisions on prices, their own experience and – depending on the treatment – grades and investment decisions. I observe that buyers are loyal if they earned a relatively high payoff in the previous round and tend to switch to another seller if their previous payoff was relatively low. Thus, the average payoff preceding a switch is significantly lower than the average payoff followed by the decision to purchase from the same seller again ($p < 0.001$). Hence, they punish sellers when they are unsatisfied with their payoff although they cannot always be sure whether a low payoff resulted from a selfish recommendation or whether they did not have a higher valuation for any version traded in the previous round.

Fig. 5 illustrates the share of buyers purchasing from the cheapest seller. In all treatments the majority of buyers purchases from the seller(s) setting the lowest price. In the baseline treatment, buyers decide in favor of the cheapest seller in almost 90% of all purchase decisions. When watchdogs are present or investment decisions are observable, buyers seek

[17] As only informed sellers can become aware of this conflict of interest, the following analysis deals only with sellers who have invested.

Table 4
Regression: share of consumer surplus achievable by following the recommendation.

Conflict intensity	−0.479***
	(0.049)
Watchdogs	0.051
	(0.035)
Watchdogs * Conflict intensity	0.135**
	(0.065)
Observability	−0.016
	(0.030)
Period	−0.002
	(0.010)
Period^2	−0.000
	(0.000)
Constant	0.922***
	(0.083)
R^2	0.242
N	739

Notes: Dependent variable is the share of consumer surplus achievable by following the advice. Output from a random effects panel regression. Cluster-robust standard errors in parentheses (on subject-level). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. When clustering on matching-group, the interaction term (Watchdogs*Conflict intensity) loses its significance (p = 0.124), while the Watchdogs-dummy is now slightly closer to significance (p = 0.107 instead of 0.143).

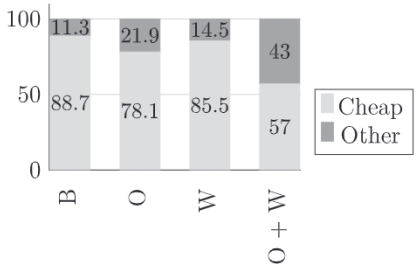


Fig. 5. Share of buyers purchasing from the cheapest seller.

the seller with the lowest price in 85% and 78% of the purchase decisions, respectively. For the treatment with watchdogs only, the null hypothesis of no treatment effect compared to the baseline treatment cannot be rejected. In the treatments with transparent investment decision, however, buyers are significantly less price sensitive (p < 0.001 when watchdogs are present and p = 0.076 in the absence of watchdogs). The significant impact of the observable investment decision on price-sensitivity is remarkable as it does not alter sellers' incentives to give good recommendations.

When at least one seller has invested and buyers can observe the outcome of the investment decision, the majority of buyers selects a seller who has invested. In the treatment without watchdogs, 56% of buyers interact with an informed seller when given the choice, whereas in the 'O+W' treatment, the preference for an informed seller is more pronounced with 68%. Hence, watchdogs seem to increase the attractiveness of informed sellers, possibly because the expected quality of their recommendations is superior.

To investigate buyers' purchase decisions more closely, I perform a regression analysis of buyers' propensity to interact with an informed seller(see Table 5). Note that this part of the analysis only considers the treatments with observable investment decisions. Further, it is limited to situations in which at least one seller in a market has invested. The explanatory variable $\text{Price}_{\text{Informed}}$ denotes the average price set by informed sellers, while the variable $\text{Price}_{\text{Uninformed}}$ denotes the average price of uninformed sellers in a market. It makes intuitive sense that buyers are more likely to select an informed seller if prices charged by uninformed sellers are higher and prices of informed sellers lower. Interestingly, they are more likely to choose an informed seller in the presence of watchdogs showing that they trust in the positive impact of watchdogs on informed sellers' recommendations. Furthermore, a small non-linear negative time-trend can be observed showing that more experienced buyers learn that recommendations from informed sellers are not necessarily better than uninformed recommendations.

Buyers cannot tell whether the *current* round is a test round, but they are informed if watchdogs have been present in the past and which grades have been awarded in their market. One could imagine that buyers react to this information by seeking sellers with good grades as they might expect these sellers to continue giving good recommendations. Fig. 6 depicts the average number of buyers that sellers were able to attract in the test round and in the four rounds after the appearance of watchdogs. For the figure, I combine all observations on test rounds and set the test round equal to zero. Moreover, I only consider markets in which different grades were awarded. As there were no markets in which a '1', a '3' and a '5' were awarded in the same round, I combine all three possible combinations of grades in the figure by referring to the relatively

Table 5
Regression: probability of choosing informed seller.

$Price_{Informed}$	−0.110***
	(0.013)
$Price_{Uninformed}$	0.045***
	(0.010)
Watchdogs	0.142*
	(0.079)
Period	−0.042***
	(0.014)
$Period^2$	0.001**
	(0.000)
Constant	1.406***
	(0.178)
R^2	0.140
N	948

Notes: Dependent variable is a dummy indicating whether an informed seller is chosen. Output from a random effects panel regression. Cluster-robust standard errors in parentheses (on subject-level). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Results remain unchanged when clustering on matching-group-level.

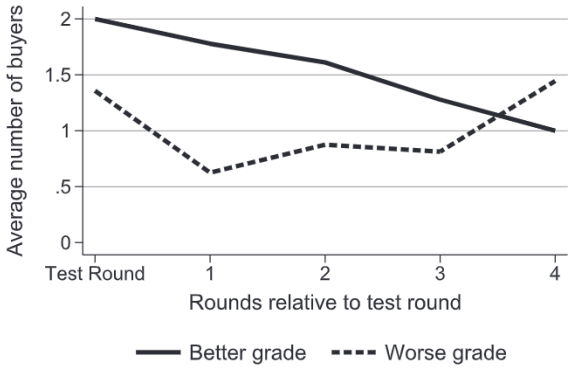


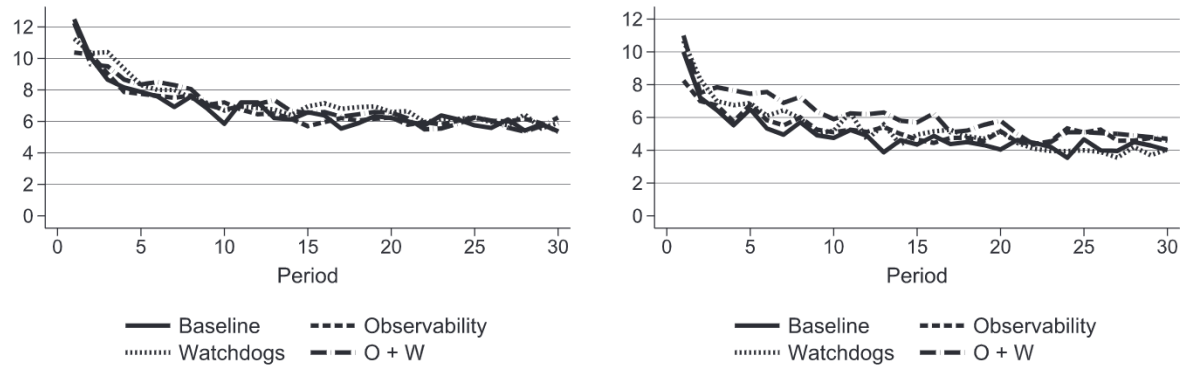
Fig. 6. Average number of buyers per seller relative to the awarded grade.

better and worse grade. From Fig. 6 it becomes clear that the above-mentioned reasoning is not widespread among buyers: Even in the first round following the test round sellers with better grades are not chosen by significantly more buyers than their competitors with worse grades ($p = 0.449$) and also afterwards buyers do not avoid sellers with bad grades. The low price level which ensures a relatively high consumer surplus even when random purchase decisions are made might be an explanation for this observation. Hence, building a reputation is not very attractive as buyers do not honor the effort of past rounds.

5.3. Prices

At the beginning of a session, prices in all treatments are high, sometimes even exceeding a price of ten which would result in an equal split of the expected payoff. Gaining more experience, sellers realize that lower prices attract more buyers and adjust their prices accordingly (see Fig. 7 a). After about ten rounds, posted prices stabilize at six ECUs and are thus higher than predicted. As can be seen from the graph, posted prices do not differ across treatments. For transaction prices (see Fig. 7 b), i.e. prices at which trade has actually taken place, a very similar pattern can be observed at a slightly lower level as buyers are price-sensitive. In the treatments with unobservable investment decisions transaction prices converge to 4 ECUs, whereas trade takes place at slightly higher prices in the treatments with observable investment decisions. From the graph it becomes clear that this is mainly driven by the higher transaction prices in the 'O+W' treatment. However, the difference to the treatment with watchdogs, but concealed investment decisions is insignificant ($p = 0.347$).

Although investment costs are sunk, sellers might try to pass them on to buyers. They might also try to signal their investment decision with a higher price when the investment decision is not observable. Regressions 1 and 2 in Table 6 show that, regardless of the observability of the investment decision, informed sellers do not charge higher prices than their uninformed competitors. Hence, in line with the behavioral predictions, there is no evidence that sellers try to pass on their investment costs to buyers or signal their investment decision through higher prices. In fact, the only explanatory variable with a significant influence on posted prices is the period number. For transaction prices, in contrast, I observe a significant and positive effect of sellers' investment decisions in the treatments with observable investment decisions (Regression 4). Hence, buyers are less price-sensitive towards informed sellers. Neither for posted nor for transaction prices



(a) Posted prices (b) Transaction prices

Fig. 7. Time series of prices.

Table 6
Regressions: posted and transaction prices.

	Posted prices		Transaction prices	
	No observability (1)	Observability (2)	No observability (3)	Observability (4)
Investment^{+}	0.266	0.090	−0.037	0.802***
	(0.442)	(0.253)	(0.170)	(0.164)
Watchdogs	0.409	0.179	0.196	0.483
	(0.554)	(0.411)	(0.267)	(0.339)
Period	−0.135***	−0.255***	−0.146***	−0.180***
	(0.050)	(0.050)	(0.027)	(0.032)
Period^2	0.002	0.005***	0.002***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Constant	8.001***	9.114***	6.315***	6.623***
	(0.405)	(0.548)	(0.292)	(0.453)
R^2	0.046	0.083	0.132	0.132
N	1076	1080	1057	1055

Notes: Dependent variables are posted prices (1 and 2) and transaction prices (3 and 4). Output from random effects panel regressions. Cluster-robust standard errors in parentheses (on subject-level). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Results remain unchanged when clustering on matching-group-level. $^{+}$For buyers, investment refers to their interaction partner's investment.

Table 7
Investment decisions.

Setting	No watchdogs	Watchdogs	p-value
No observability	15.3%	17.3%	*0.583*
Observability	40.3%	51.0%	*0.273*
p-value	*0.030*	*0.028*	

does the presence of watchdogs have a significant impact. Both findings contradict the behavioral predictions: Informed sellers do not have an incentive to give better recommendations while watchdogs contribute to closing the information asymmetry between buyers and sellers and might thus improve the quality of recommendations justifying higher prices.

5.4. Investment decisions

Investing seems unattractive for sellers as they cannot pass on their costs to buyers. Nevertheless, at the beginning of a session the majority of sellers invests regardless of the treatment (see Fig. 8). Growing more experienced, sellers cease to invest, but the investment rates do not decrease to 0 in any treatment. Over time, the increasing effect of the observability of investment decisions on the share of investing sellers becomes visible.

Table 7 presents the average share of informed sellers for each treatment. For both implementations of investment decisions – unobservable and observable – watchdogs do not influence investment behavior. Pooling over the treatments with and without watchdogs, the impact of revealing investment decisions is highly significant ($p = 0.002$): When buyers know which sellers have invested, sellers are more likely to invest. This observation stays significant when analyzing the treat-

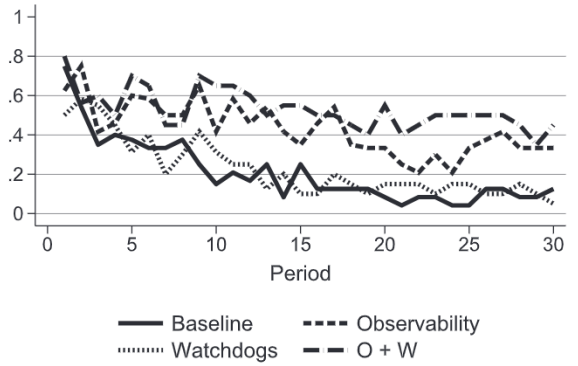


Fig. 8. Time series of investment decisions.

Table 8
Regressions: number of clients.

	(1) No observability	(2) Observability
Price	−1.077***	−0.800***
	(0.095)	(0.077)
Price^2	0.048***	0.037***
	(0.006)	(0.004)
Investment$^+$	−0.007	1.043***
	(0.429)	(0.226)
Price * Investment	−0.007	−0.106***
	(0.050)	(0.027)
Watchdogs	0.167	0.092
	(0.122)	(0.181)
Period	−0.029***	−0.022***
	(0.006)	(0.006)
Constant	6.066***	4.622***
	(0.440)	(0.371)
R^2	0.392	0.238
N	1076	1080

Notes: Dependent variable is the number of buyers a seller could attract. Output from random effects panel regressions. Cluster-robust standard errors in parentheses (on subject-level). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Results remain unchanged when clustering on matching-group-level. $^+$For buyers, investment refers to their interaction partner's investment.

Table 9
Mean payoffs.

	Baseline	O	W	O + W
Sellers	0.91	0.85	0.68	1.33
Buyers	10.63	10.19	10.66	10.31
Total	11.54	11.04	11.34	11.64

ments with watchdogs and without watchdogs separately and cannot be reconciled with the behavioral predictions based on the rationality assumption.

Although sellers do not demand higher prices when they have invested, they might still find it profitable to invest if it helps them attract more clients. I therefore regress the number of attracted buyers on the seller's investment decision while controlling for prices, experience and the presence of watchdogs (see Table 8). Sellers who have invested indeed attract significantly more buyers than their uninformed competitors in the treatments with observable investment decisions. Both with observable and unobservable investment decisions do prices have a decreasing, but non-linear effect on the number of clients. I find a negative coefficient for the interaction of price and investment decision showing that consumers are also very price-sensitive towards informed sellers.

5.5. Distribution of rents

I now proceed to analyze whether the introduction of watchdogs and the revelation of investment decisions lead to differences in the distribution of rents. As can be seen in Table 9, sellers' payoffs in all treatments are on average very low due to intense price competition.

Table 10
Regressions: payoffs.

	(1) Sellers	(2) Buyers
Price	0.825***	−0.554***
	(0.255)	(0.215)
Price^2	−0.052***	−0.037**
	(0.015)	(0.019)
Investment$^+$	−1.780***	0.424
	(0.365)	(0.311)
Observability	−0.013	−0.212
	(0.314)	(0.202)
Investment * Observability	1.621***	0.538
	(0.539)	(0.395)
Watchdogs	0.096	0.395**
	(0.361)	(0.159)
Period	−0.069***	−0.018**
	(0.016)	(0.009)
Constant	−0.563	14.177***
	(1.263)	(0.654)
R^2	0.051	0.200
N	2156	2112

Notes: Dependent variable is the payoff in a round. Output from random effects panel regressions. Cluster-robust standard errors in parentheses (on subject-level). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. When clustering on matching-group-level, the Watchdogs-dummy loses its significance ($p = 0.102$ instead of 0.013). $^+$For buyers, investment refers to their interaction partner's investment.

Buyers benefit from the fierce price competition in all four treatments by paying low prices. The treatment with observable investment decisions and watchdogs, intended to make buyers better off, actually maximizes *sellers'* surplus. Although more sellers invest and incur investment costs, they can recover these costs and even enhance their surplus due to the slight increase in transaction prices. On the other market side, the match between buyers' preferences and purchased versions is improved due to better recommendations, yet the increase in transaction prices offsets this positive effect on buyers' surplus. Note that neither for sellers nor for buyers are any of the differences in payoffs between the treatments statistically significant.

The observation of payoff equality with and without regulatory measures follows also from Baron (2011)'s model of a vertically differentiated credence goods market. When the standard of the credence characteristic is increased, producers benefit from the consumers' higher willingness to pay. The welfare effects due to the increase in transaction prices and the improvement of the credence characteristic are identical in size.

To complement the results above, I perform a regression analysis of the profits controlling for prices, experience and interaction effects. As can be seen from Table 10, sellers suffer from investing when their decision is not revealed, yet in the treatments with observable investment decisions, their payoff is not significantly increased by their choice to invest. Regardless of the observability of investments, interacting with an informed seller does not have a significant impact on buyers' payoff. Here, it is the presence of watchdogs which increases buyers' payoff by improving the quality of sellers' recommendations. Prices have a non-linear effect on sellers' payoff: Sellers benefit from small increases due to higher earnings per transaction, but suffer from larger increases because of a reduction in the number of transactions.

6. Discussion and conclusion

This article investigates the impact of watchdogs and observable investment decisions on market outcomes in horizontally differentiated credence goods markets. In these markets buyers cannot identify the product version appropriate for their needs before purchasing. Although they experience their utility from consumption, they do not find out if they have purchased their personally optimal variant ex post. Thus, in terms of the available information, the considered product category is situated between experience goods and label credence goods (as described, for example, in Roe and Sheldon, 2007; Baksi and Bose, 2007; Feddersen and Gilligan, 2001): Consumers do experience their utility from consumption – a feature which differentiates the products considered in this paper from label credence goods, where main product characteristics remain unobservable ex post. Yet, in contrast to experience goods, lay consumers cannot fully judge their purchase decision ex post because in many situations they do not know if a different product would have been a better match for their preferences. In this sense, the goods correspond to the definition of credence goods in the literature (see e.g. Dulleck et al., 2011; Balafoutas and Kerschbamer, 2020). Hence, the market considered in this article features a smaller information asymmetry than a market for label credence goods, but a larger information asymmetry than a market for experience goods. This information asymmetry is detrimental for buyers, as it leads to suboptimal matches between their preferences and purchases.

While the boundary between the type of credence goods considered here and label credence goods is relatively clear-cut, the degree of differentiation with respect to experience goods depends on the level of the consumers' expertise outside

the lab. If consumers are able to infer the fit between their preferences and their purchase from the experienced utility, the goods qualify as experience goods. Lay consumers, in contrast, lack the knowledge to make this inference, i.e. to them such products constitute credence goods. If all consumers were experts in every area of their consumption, there would indeed be no need for the institutions analyzed here – watchdogs and observable investment decisions. Yet, as consumers need to make a plethora of purchase decisions, it is unlikely that they are sufficiently knowledgeable to competently judge every purchase decision ex post. To represent this situation in the laboratory, parameters are chosen such that subjects lack the information necessary to judge the fit between their preferences and their purchases. Up to the present, this type of credence good has mainly been investigated in the context of vertical product differentiation. By use of an experiment, I seek to understand the functioning of credence goods markets with a horizontal product differentiation component and assess the impact of potentially helpful information-providing institutions.

In the experiment, subjects trade differentiated goods on markets taking the roles of buyers and sellers. Each seller offers the same product variants as his competitors, but he features different profit margins for each variant. Consumers differ in their valuations for the product versions which are unknown to them. Consequently, they cannot identify their most preferred variant. Sellers can recommend a variant to their customers. However, they need to invest, which is costly, in order to be able to observe their clients' valuations. Otherwise, they only know their own costs for each variant which do not include any information on consumer preferences. Observing a recommendation, consumers do not know if their seller has invested and is thus *capable* of giving a useful recommendation. Even if they assume that their seller has invested, they do not know if their seller is *willing* to take his knowledge about their preferences into account when deciding which version to recommend. Having followed his recommendation, they cannot judge its quality: Observing their payoff, they can derive their valuation for the purchased variant. Yet, they do not know for sure if they would have had a higher valuation for another variant. As consumers will almost never find out if their sellers have given reliable recommendations, sellers cannot build a reputation of giving good recommendations.

To improve the match between preferences and purchases in horizontally differentiated credence goods markets, two information-providing institutions are implemented: The sellers' investment decisions are revealed and watchdogs are introduced which test the quality of randomly selected recommendations. Watchdogs are designed to resemble checks by consumer organizations whereas the revelation of the sellers' investment decisions comes close to the public display of qualification certificates. These information-providing institutions are particularly attractive, as the latter appears to be easily implementable and the former may even evolve without regulatory interventions and sustain with low or no subsidies.

In the baseline treatment with hidden investment decisions and no checks of the recommendation quality, the results are in line with the behavioral predictions based on the assumption of rational payoff-maximization on both market sides. Very few sellers invest and the uninformed sellers give recommendations which minimize their costs. Even those who have invested give selfish recommendations. The price level is low as consumers are very price-sensitive. Although they might anticipate that their interaction partner has not invested and gives selfish advice, they still follow about half of his recommendations – a behavior which does not reduce their expected payoff as their valuations and sellers' costs are uncorrelated. When the investment decision is revealed, more sellers decide to invest. Indeed, they rather invest to attract consumers than to base their recommendations on the acquired information as the share of selfish recommendations given by informed sellers is unchanged compared to the baseline treatment. Knowing that each recommendation may be subject to a check by watchdogs, informed sellers give more consumer-friendly recommendations. The share of informed sellers is yet unchanged in the watchdogs treatment compared to the baseline treatment. Buyers are very price-sensitive in this treatment as well and react little to grades. If both investment decisions are revealed and watchdogs appear from time to time, a substantial fraction of sellers invests and gives consumer-friendly recommendations. Buyers accept higher prices such that sellers who have invested do not incur losses in expectation. Also the fraction of buyers following sellers' recommendations is highest in this treatment. Hence, with both observable investment decisions and watchdogs matches between choices and preferences are improved compared to the other treatments.

The situation in the baseline treatment is comparable to a market in which only discounters exist. Although they do not receive reliable advice, low prices ensure that consumers benefit from their purchases even if they happen to purchase their least preferred variant. The revelation of investment decisions induces more sellers to invest, i.e. the share of expert sellers relative to discounters increases, but the quality of advice is unchanged and so is the match between consumers' preferences and purchased variants. Watchdogs improve the quality of recommendations of expert sellers, yet most sellers opt against becoming experts. Only when investment decisions are revealed *and* watchdogs are present, a substantial share of expert sellers is active on the market and incorporates their knowledge about buyers' preferences into their advice.

The treatment variations have no significant effect on the distribution of welfare: In the case of buyers, their willingness to purchase at higher prices offsets the welfare improvement due to better matches between choices and preferences. For sellers, the increase in transaction prices compensates their investment expenditures. Hence, the simultaneous introduction of watchdogs and the revelation of sellers' investment decisions provokes a switch from a discounter market with low prices and low-quality recommendations to a market with higher transaction prices and informative product recommendations, yet in terms of welfare, market constellations do not differ. Note that the welfare-equality of both market constellations is due to the design of the investment costs in the experiment which are calibrated to ensure welfare-neutrality of investments. One can conjecture that, with lower investment costs, an improvement of overall welfare can be reached by jointly introducing both information-providing institutions.

Table A1
Probabilities.

Valuation	10 – 13	14	15	16	17	18	19	20
Ex ante	0%	0.22%	1.1%	3.2%	7.58%	15.15%	27.27%	45.45%
Conditional	0%	0.22%	1.12%	3.24%	7.9%	17%	37.48%	100%

Interestingly, two policy measures which – by themselves – proved ineffective in improving the match between preferences and purchases have a significant impact when jointly introduced. One of the measures – the revelation of the investment decision – does provide information, but does not influence the incentive structure such that its introduction should remain inconsequential, yet it increases the share of expert sellers. The introduction of watchdogs, however, may have changed investment behavior if buyers and sellers were rational, instead, it only increased the consumer-friendliness of the very few informed recommendations. Due to the bounded rationality of decision makers, the two measures can only function in combination. In fact, the boundedly rational reaction to the revelation of investment decisions is necessary to ensure that watchdogs can function. Hence, when designing policy measures decision makers' deviations from rationality need to be considered and measures need to be jointly introduced, of which some, under the assumption of rational market participants, would not have been necessary. In order to promote expert-markets for horizontally differentiated credence goods, governments should require retailers to display qualification certificates and ensure, possibly through subsidies, that consumer organizations perform random tests.

Of course, behavior in the experiment depends on specific design choices and it may differ if the implementation of the investment decision was changed, i.e. if sellers invest knowing how many buyers want to interact with them and investment is specific to the buyer. While I conjecture that the latter change should only have a minor impact on behavior as each seller attracts on average one buyer under the current design, the former might induce a larger change in behavior: Knowing that the buyer has trusted and therefore selected them, sellers might feel more inclined to reciprocate trust, invest and give a good recommendation. Besides, this change in timing prevents the sellers from making losses – a fear which might have prevented sellers from investing under the current design. However, as the main results are derived from between-treatment comparisons, I am confident that they are robust to a change in the design of the investment decision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

Appendix A contains a calculation of the probabilities.

Probabilities. Costs and valuations are drawn without replacement from a discrete uniform distribution, $\{0, 1, 2, \ldots, 10\}$ for sellers and $\{10, 11, 12, \ldots, 20\}$ for buyers, respectively. There are five draws for each subject in each round. Thus, the probability of each cost or valuation sample equals $\frac{1}{11}\frac{1}{10}\frac{1}{9}\frac{1}{8}\frac{1}{7} = \frac{1}{55440}$. As the order of the draws does not matter, the probability is multiplied by 5! such that each sample – independent of the order of the draws – occurs with a probability of 0.21645%. Let us consider the buyers' valuations and compute the probability that a certain valuation is the highest valuation in a draw: Values below 14 cannot be the highest draw in a sample as there are five independent draws without replacement. 14 occurs with a probability of 0.21645% as the highest draw because there is only one possible combination of numbers with 14 as the highest draw, i.e. there is only one path leading to this sample. For valuations above 14 the probability needs to be multiplied by the number of possible paths leading to samples with the same highest valuation. For a highest valuation of 15, there are five slots k to fill with numbers between 10 and 15. Hence, there are six different numbers n that can potentially be part of the sample. The number of samples in which 15 is not the highest valuation needs to be subtracted, such that $\frac{n!}{k!(n-k)!} - \frac{(n-1)!}{k!(n-k-1)!}$. Still only focusing on draws between 10 and 15, there is one possible sample involving the numbers from 10 to 14 and not including 15. We need to take this sample into account when computing the probability that 15 is the highest valuation in a sample: $0.21645\%\left(\frac{6!}{5!1!} - \frac{5!}{5!0!}\right) = 1.1\%$.

Based on the ex ante probability of observing a sample with a certain highest valuation, the probability that a given valuation is the highest in a sample can be derived. Hence, consumers can compute the probability that an implemented recommendation has been honest. The rationale behind the computation is as follows: Observing a certain valuation all samples consisting only of lower valuations can of course be excluded from the set of possible valuation samples. The probabilities that an observed valuation is the highest in a sample and the ex ante probabilities are summarized in Table A.1.

Appendix B

Appendix B contains the instructions for the baseline treatment.

Instructions. Thank you for participating in this experiment. Please do not talk to the other participants during the experiment.

This experiment consists of 30 rounds. There are two different roles you can take: seller and buyer. Your role is determined randomly at the beginning of the experiment and you keep your role throughout the whole experiment. Which role you take will be displayed on your screen.

The participants are split into two groups consisting of four sellers and four buyers each. These groups persist for the whole experiment. As a buyer, your potential interaction partners are Seller 1, Seller 2, Seller 3 and Seller 4. If you are a seller, your potential interaction partners are Buyer 1, Buyer 2, Buyer 3 and Buyer 4. Note that the numbering of the sellers is fixed, i.e. the same number always describes the same participant. The numbering of the buyers is not fixed and changes from round to round.

In each round sellers offer the same five product variants for which different costs occur at the different sellers. These costs always lie between 0 and 10 ECUs. The sellers own enough units of each variant to serve all buyers in their group. Products and costs vary from round to round. Buyers have different valuations for the different variants. These valuations, which are randomly determined in each round and lie between 10 and 20 ECUs, are unknown to buyers. For this reason sellers have the possibility to give advice and assist buyers in choosing a variant. In order to know the buyers' valuations, a seller needs to invest in the beginning of a round. Investing costs 1.5 ECUs. A seller can only observe the buyers valuations and give useful recommendations if he has invested. Buyers do not know if their seller has invested and if the recommendation is honest, i.e. if the seller has recommended the variant for which they have the highest valuation.

Payoffs in a round

Buyers

If a buyer has bought a product variant, her payoff equals her valuation for the chosen variant minus the paid price. If a buyer has not bought anything, her payoff in this round equals zero.

Sellers

If a seller has found *one* buyer, his payoff equals his price minus the costs of the sold variant minus his investment costs if he has invested.

If he has attracted *several* buyers, his payoff will be the sum of the payoffs from the individual transactions. The investment costs will only be subtracted once if he has invested.

If a seller has not attracted a buyer and he has not invested, his payoff equals zero. If he has invested but no buyer has purchased from him, his payoff equals -1.5 ECUs.

Timing of a round

1. Investment decision: Sellers decide if they want to invest. For the investment, costs of 1.5 ECUs occur.
2. Price setting: Each seller sets one price for his five product variants. This price needs to lie between 1 and 20 ECUs and applies to all five product variants. Each seller offers the same five product variants. When deciding on his price, the seller observes his costs for the five product variants. These costs occur when a variant is sold and they are independent of the investment costs. Note: If a seller does not set a price, the computer will save a price of 20 ECUs after 35 seconds. Thus, he will most likely not attract a buyer in this round.
3. Selection of a seller: Buyers observe the prices set by Seller 1, 2, 3 and 4 in their group. Buyers cannot identify the seller(s) who invested. Each buyer decides with which seller she wants to interact. If she does not select a seller, the round ends for her. As a reminder her past transactions are displayed on her screen.
4. Advice: Sellers get to know which buyers have chosen to interact with them. At most, a seller can attract all four buyers from his group. If the seller has invested, he can observe the buyers' valuations for the different product variants. The valuations differ among the buyers. For each interacting buyer the net utility (= valuation - price) of each variant is displayed. The seller also observes his margin (= price - costs) for each product variant. If he has not invested, he only observes his margins. The seller decides for each buyer separately if he wants to give a recommendation and which variant to recommend.
5. Purchase: Buyers observe which variant their seller has recommended or that their seller has not given a recommendation. Buyers can follow the advice and purchase the recommended variant, but they can also choose another variant or decide not to purchase in the current round.
6. Payoff: The earnings are displayed on the screens. Sellers are informed if they could sell something and which variants their buyers have chosen. If they have purchased, buyers are informed about their net utility (= valuation - price) of the purchase.
7. New round: Buyers receive new identification numbers. In the next round different product variants will be traded, i.e. the sellers' costs and the buyers' valuations will again be determined randomly. The valuations always lie between 10 and 20 ECUs and the costs between 0 and 10 ECUs, but the actual values change.

At the end of the experiment 10 of the 30 rounds are randomly selected for payment. Only these ten rounds are payoff relevant. Your average income of these 10 rounds will be converted into Euros. For sellers the exchange rate equals ECUs * 5 = €. For buyers the exchange rate is ECUs * 1 = €. In addition to their earnings from the selected 10 rounds sellers receive € 5 and buyers € 2.50 as a fixed payment. The earnings are paid out privately and in cash at the end of the experiment.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jebo.2020.12.018.

References

Anbarcı, N., Feltovich, N., Gürdal, M.Y., 2015. Lying about the price? ultimatum bargaining with messages and imperfectly observed offers. J. Econ. Behav. Org. 116, 346–360.
Baksi, S., Bose, P., 2007. Credence goods, efficient labelling policies, and regulatory enforcement. Environ. Resour. Econ. 37 (2), 411–430.
Balafoutas, L., Beck, A., Kerschbamer, R., Sutter, M., 2013. What drives taxi drivers? a field experiment on fraud in a market for credence goods. Rev. Econ. Stud. 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. J. Behav. Exp. Financ. 26, 100285.
Baron, D.P., 2011. Credence attributes, voluntary organizations, and social pressure. J. Public Econ. 95 (11–12), 1331–1338.
Bonroy, O., Constantatos, C., 2008. On the use of labels in credence goods markets. J. Regul. Econ. 33 (3), 237–252.
Conrads, J., Irlenbusch, B., 2013. Strategic ignorance in ultimatum bargaining. J. Econ. Behav. Org. 92, 104–115.
Corazzini, L., Kube, S., Maréchal, M.A., Nicolo, A., 2014. Elections and deceptions: an experimental study on the behavioral effects of democracy. Am. J. Pol. Sci. 58 (3), 579–592.
Dana, J., Weber, R.A., Kuang, J.X., 2007. Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. Econ. Theory 33 (1), 67–80.
Darby, M.R., Karni, E., 1973. Free competition and the optimal amount of fraud. J. Law Econ. 16 (1), 67–88.
Dulleck, U., Johnston, D., Kerschbamer, R., Sutter, M., 2012. The good, the bad and the naive: do fair prices signal good types or do they induce good behaviour? Working Papers in Economics and Statistics 2012-03. University of Innsbruck.
Dulleck, U., Kerschbamer, R., 2006. On doctors, mechanics, and computer specialists: the economics of credence goods. J. Econ. Lit. 44 (1), 5–42.
Dulleck, U., Kerschbamer, R., 2009. Experts vs. discounters: consumer free-riding and experts withholding advice in markets for credence goods. Int. J. Ind Org. 27 (1), 15–23.
Dulleck, U., Kerschbamer, R., Sutter, M., 2011. The economics of credence goods: an experiment on the role of liability, verifiability, reputation, and competition. Am. Econ. Rev. 101 (2), 526–555.
Emons, W., 1997. Credence goods and fraudulent experts. Rand. J. Econ. 107–119.
Feddersen, T.J., Gilligan, T.W., 2001. Saints and markets: activists and the supply of credence goods. J. Econ. Manag. Strat. 10 (1), 149–171.
Fischbacher, U., 2007. Z-tree: zurich toolbox for ready-made economic experiments. Exp. Econ. 10 (2), 171–178.
Gehrig, T., Güth, W., Levati, V., Levinsky, R., Ockenfels, A., Uske, T., Weiland, T., 2007. Buying a pig in a poke: an experimental study of unconditional veto power. J. Econ. Psychol. 28 (6), 692–703.
Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with orsee. J. Econ. Sci. Assoc. 1 (1), 114–125.
Güth, W., Schmittberger, R., Schwarze, B., 1982. An experimental analysis of ultimatum bargaining. J. Econ. Behav. Org. 3 (4), 367–388.
Huck, S., Lünser, G.K., Tyran, J.-R., 2012. Competition fosters trust. Games Econ Behav 76 (1), 195–209.
Huck, S., Lünser, G.K., Tyran, J.-R., 2016. Price competition and reputation in markets for experience goods: an experimental study. Rand. J. Econ. 47 (1), 99–117.
Kerschbamer, R., Sutter, M., 2017. The economics of credence goods–a survey of recent lab and field experiments. CESifo Econ. Stud. 63 (1), 1–23.
Kriss, P.H., Nagel, R., Weber, R.A., 2013. Implicit vs. explicit deception in ultimatum games with incomplete information. J. Econ. Behav. Org. 93, 337–346.
Mimra, W., Rasch, A., Waibel, C., 2016a. Price competition and reputation in credence goods markets: experimental evidence. Games Econ. Behav. 100, 337–352.
Mimra, W., Rasch, A., Waibel, C., 2016b. Second opinions in markets for expert services: experimental evidence. J. Econ. Behav. Org. 131, 106–125.
Mitzkewitz, M., Nagel, R., 1993. Experimental results on ultimatum games with incomplete information. Int. J. Game Theory 22 (2), 171–198.
Nelson, P., 1970. Information and consumer behavior. J. Polit. Econ. 78 (2), 311–329.
Pesendorfer, W., Wolinsky, A., 2003. Second opinions and price competition: inefficiency in the market for expert advice. Rev. Econ. Stud. 70 (2), 417–437.
Roe, B., Sheldon, I., 2007. Credence good labeling: the efficiency and distributional implications of several policy approaches. Am. J. Agric. Econ. 89 (4), 1020–1033.
Schneider, H.S., 2012. Agency problems and reputation in expert services: evidence from auto repair. J. Ind. Econ. 60 (3), 406–433.