CESifo Economic Studies, 2017, 1–23
doi: 10.1093/cesifo/ifx001
Advance Access Publication Date: 13 February 2017
Original article

The Economics of Credence Goods – a Survey of Recent Lab and Field Experiments*

Rudolf Kerschbamer[#] and Matthias Sutter[§]

[#]University of Innsbruck and [§]University of Cologne, University of Innsbruck, CESifo Munich and IZA Bonn

Abstract

Many important markets, such as those for health care or repair services, are characterized by pronounced informational asymmetries between buyers and sellers – because the buyer cannot identify the quality of the product that fits her needs best, while the seller can do so by performing a diagnosis. The seller can then recommend the appropriate or an inappropriate quality. Since the buyer has no means to verify the correctness of the recommendation but has to rely on the honesty of the seller, such goods are called credence goods. Markets for credence goods are typically plagued by incentives for fraudulent behavior and by inefficiencies caused by actual or expected fraud, covering overprovision and overcharging. Although highly policy relevant, a carefully controlled analysis of the provision of credence goods through the use of experiments has only recently taken off. Here, we provide an overview about recent laboratory and field experiments on the economics of credence goods and discuss important implications and directions for future research.

JEL classification: C91, C93, D82

Key words: credence goods, asymmetric information, policy, lab experiment, field experiment

1. Introduction

In many important markets – as in those for health care, repair and legal services, as well as in those for financial advice and fund management – consumers (patients, clients or private investors) are unable to identify the quality of a good, service or asset that fits their needs best. They may even be unable to verify the quality that they have actually received.

* This paper is based on Sutter's keynote at the CESifo area conference on employment and social protection, 16 May 2015, in Munich. We acknowledge helpful comments from Kai Konrad, Panu Poutvaraa, two referees and conference participants. Financial support from the Austrian Science Fund (FWF grants P20796, P26901 and special research area grant SFB 63) and from the Australian Research Council (ARC Grant DP110103653) is gratefully acknowledged.

In contrast, doctors, mechanics and legal or financial experts are typically better informed regarding the appropriate quality of service provision. As this information asymmetry between trading partners often persists even after a trade has been concluded and the buyer has consumed the good or service, such goods and services are referred to as credence goods (see e.g., Darby and Karni, 1973, Dulleck and Kerschbamer, 2006, Huck et al., 2016a).[1]

The volume of trade on credence goods markets is huge. For instance, health care expenditures account for about 10% of GDP in the OECD-countries alone (www.oecd-library.org). A significant portion of these expenditures is caused by the provision of medical treatments where the prescribing physicians have a large informational advantage over their patients who might not only be uninformed about the most efficient type of treatment, but who may even ex post be unable to distinguish a cheap from an expensive drug infusion. Repair services are also a multi-billion dollar industry. In the EU, car repairs alone are worth about 100 billion Euro per annum (ec.europa.eu/eurostat), with a significant proportion of car repairs being unnecessary (Hubbard, 1998), probably in part because mechanics exploit their superior information about the appropriate service. Moreover, ex post inspection by the customer may fail to distinguish a replaced part from a repaired part. Also, the finance sector is one of the biggest industries worldwide. Its share of GDP has increased over the past 50 years in major Western economies from an average of 3% to about 9%. The complexity of its products has created severe informational asymmetries between advisors and clients. Moreover, the inherent conflict between two tasks performed by financial advisors – prospecting for customers and advising on the product's "suitability" for the specific needs of customers – has created misaligned incentives (Inderst and Ottaviani, 2012). This implies that clients are exposed to potentially malign behavior of financial professionals in the form of unsuitable product provision.

In general, the informational asymmetries between expert sellers and their customers on any market for credence goods create strong material incentives for misbehavior on the side of the seller. If not contained by institutional remedies or moral constraints these misaligned incentives translate into large efficiency costs for society as a whole. For this reason, it seems important to investigate the provision of credence goods to get a better understanding of the determinants of misbehavior of sellers and of the factors that can promote a more efficient provision of credence goods.

To define the various types of fraudulent behavior on markets for credence goods more systematically, consider a car owner bringing her vehicle to a garage for repair. The mechanic might have an incentive to cheat the consumer on two levels: First, the repair might be inefficient due to the mechanic replacing more parts than are actually necessary to bring the car back on the road (and charge for the additional time and material). This case is referred to as *overprovision* because the additional costs of the high quality service are larger than the benefits to the consumer. The mechanic's repair might also be insufficient, thus leaving the consumer with a bill, but with a car that is still not running properly. This latter case is

1 It is important to note that there exist two more or less disjunct strands of literature on credence goods. For one strand the defining characteristic of a credence good is that the consumer is unable to identify the quality that fits her needs best. This is the strand of literature we are referring to in this article. The second strand of literature takes the characteristic that credence goods have qualities which are expensive to judge even after purchase or concumption as the defining feature of a credence good. Typical examples mentioned in this second strand are goods vertically differentiated by process-attributes (such as, e.g., whether food has been produced organically or not).

referred to as *underprovision* since any material and time spent on the repair is a pure waste. Second, the repair might be appropriate, but the mechanic might charge the consumer for more than he has actually done. This kind of problem is known as *overcharging* and it can also lead to inefficiencies in the long run if the fear of getting overcharged deters consumers from trading on credence goods markets in the future. Such a process might ultimately lead to a type of market-breakdown as in Akerlof (1970).

Several *theoretical contributions* have addressed the conditions for the efficient – and honest – provision of credence goods (e.g., Darby and Karni, 1973, Wolinsky, 1993, Sülzle and Wambach, 2005). A survey by Dulleck and Kerschbamer (2006) organizes the different assumptions underlying these (and other) papers and provides a unifying framework. Their model shows that any of the two following conditions are sufficient for the efficient provision of credence goods: First, *liability*, meaning that the expert is required to provide a good that satisfies the customer's needs. Liability prevents the problem of underprovision, but not necessarily the problems of overprovision and overcharging. In theory, the latter problems are solved by the choice of an appropriate price structure. For this reason, policy rules that make the seller liable for providing a sufficient quality might be welcome for consumers, but also from an efficiency point of view.[2] Second, *verifiability*, meaning that the customer is able to observe and verify the type and quality of the delivered product. The classic example is to ask a car mechanic to show the replaced parts to the customer before paying the bill. Verifiability solves the problem of overcharging, but not necessarily the problems of over- and underprovision.[3] As in the liability case, these latter problems are avoided under a carefully designed price structure. In the absence of both liability and verifiability, standard theory predicts that experts will always provide the lowest quality (implying underprovision if this quality is insufficient) and charge for the highest quality (implying overcharging). If customers anticipate experts' incentives correctly, they might abstain from interaction in these markets, unless other features (like reputational concerns or competition between sellers) can make trade on these markets profitable again.

2 In practice, there are many limits to the implementation of liability rules, however. On the one hand, liability requires that the success of a particular provision of a good or service is verifiable. This is particularly difficult in the medical realm. For instance, whether some treatment has been successful to cure a patient's permanent and strong headache is difficult, to say the least, to verify for a court, but it is certainly observable by the patient. On the other hand, even in cases where success is verifiable, strict liability cannot easily be implemented. For example, after a seemingly successful repair of a car, it might still break down after a few weeks. To exclude underprovision, liability would need to cover a longer time period in such cases. Yet, during this period, the car may break down for completely unrelated reasons, or the driver may lack the required maintenance effort to prevent the break-down.

3 While liability requires somehow a verifiability of outcomes (whether the needed quality has been provided or not), verifiability requires a verifiability of inputs so that customers can check whether they actually got what sellers claim to have given them. In practice, this is also not trivial to satisfy in many situations. For instance, when showing the replaced parts after having repaired a car, sellers might simply have taken the replaced parts from other cars, yet this would be difficult to check for customers. Likewise, in medical treatment it would be extremely difficult for a patient to check whether a particular infusion contains the required chemical agents. As another example, the quality of breast implants can typically not be verified without removing them from the breasts, which is in many cases almost prohibitive.

That fraud in markets for credence goods is more than just a theoretical possibility is documented in several non-experimental *field studies* that use observational data. Probably the most important and best-studied credence goods sector is the market for health services. For this sector it is well documented that fee differentials affect provision behavior, even if medical indications would not imply that. Gruber and Owings (1996) and Gruber et al. (1999) make this point for the case of cesarean deliveries vs. natural births by showing that the relative frequency of cesarean deliveries responds to the differential remuneration for the two types of births, even though fees are unrelated to medical indications. The likelihood of cesarean deliveries also depends on whether the pregnant woman has a physician or lawyer in her family, which again is uncorrelated with what is best for the woman. However, a physician in the family arguably reduces the informational asymmetry between the doctor and the pregnant woman, and having a lawyer in the family makes it easier to enforce liability later on. Iizuka (2007) finds that Japanese doctors, who are allowed to sell pharmaceuticals at their office, tend to prescribe drugs with higher mark-ups, controlling for medical indication and drug effectiveness. Overprovision in markets for repair services is also a widespread phenomenon. For instance, in the three-digit billion dollar market for car repairs in the US, half of the repairs are estimated to be unnecessary (Hubbard, 1998).

While these non-experimental field studies impressively document the existence of fraud, they generally lack a controlled variation of variables that are predicted to influence the extent of underprovision, overprovision and overcharging. Systematic variation of factors that may affect the extent and type of fraud under *ceteris paribus* conditions is an important advantage of *experimental studies*, be they conducted in the laboratory or in the field. By carefully controlling the environment and varying single factors, experiments can identify causal effects of different conditions that are important for consumer-friendly behavior in the markets for credence goods.[4] For these reasons, the experimental analysis of credence goods has gained momentum over the past few years – as a complement to observational field studies – and it provides sound evidence on the causes of inefficiencies and fraudulent behavior in these markets. It might also pave the way for thinking about policy solutions that are intended to abate the problems that arise in these markets.[5]

In the following, we present an overview of recent experimental studies on the performance of markets for credence goods, including both lab and field experiments. As it will turn out, the currently existing body of work focuses primarily on identifying the conditions that lead to more or less fraudulent behavior of sellers. Much less is known about how different policy interventions might affect the levels of fraud. We are going to organize our overview along a few selected topics. This will be done in section 2. After that, we are going to discuss our findings and outline possible roads for future research.

4 See Friedman and Sunder (1994) for an excellent hands-on introduction to experimental methods.

5 Zitzewitz (2012) has recently introduced the term "forensic economics" for economic studies designed to detect fraud and deception (and partly also corruption) in economic activities. Randomized control trials in the field (i.e., natural field experiments) are an important means for uncovering fraudulent behavior, and in this paper we present several field experiments that examine the sources and consequences of fraud in credence goods markets. These studies clearly belong to the discipline of "forensic economics".

2. Topics in the Experimental Analysis of Credence Goods Markets

2.1 Do informational asymmetries really matter?

Despite the theoretical prediction that sellers of credence goods will exploit their informational advantage over buyers, it is not trivial to show cleanly that this is indeed the case in reality. Showing this requires a carefully controlled variation of the perception of the seller regarding the extent of the information asymmetry – or the likelihood of its presence – and the examination of the effects of this variation on the provision and charging behavior of sellers. This is what we have implemented in a field experiment in the market for taxi rides, reported in Balafoutas et al. (2013). In this experiment, we let undercover taxi passengers (i.e., experimenters) take rides to various destinations in the city of Athens, Greece. A key feature of our experimental design was that we installed so-called "triples" of rides where three experimenters took the same route at practically the same time, but acted in different roles. The first role was called "local passenger". Here, the passenger entered the taxi and only stated the requested destination. The second role was labelled "non-local native passenger". This passenger, like the local one, spoke in Greek, but when stating the destination this passenger added a phrase asking whether the taxi driver knew where the requested destination was, complementing that he was not familiar with the city. This simple addition was meant to manipulate the taxi driver's perception of the passenger's familiarity with the city such that the non-local native passenger was arguably considered less likely to be aware of the optimal, and cheapest, route to the destination than the local passenger. Finally, our third information role was that of a "foreign passenger". This passenger used the same script as the non-local native, but spoke in English. Since the content of the script was the same as for the non-local native passenger it arguably manipulated the driver's perception of the passenger's familiarity with the city in the same way. However, since taxi tariffs are nation-wide regulated and identical across Greece, the non-local native passenger was not supposed as being considered as less informed about the correct tariff – regarding day- or night-time tariffs, surcharges or the like – than the local passenger. This is different for the foreign passenger who was arguably perceived by the taxi driver as less likely knowledgeable about the details of the tariff system, for which reason the foreign passenger was expected to be cheated upon more often with respect to applying the correct tariff.

As regards the implementation, we let a triple of experimenters (each in a different information role) wait at a given taxi stand, hiding behind a corner, and then enter one after the other a taxi. All experimenters were equipped with GPS-loggers that allowed recording the exact route taken by the taxi driver. By comparing the length and time for getting to the requested destination, we could compare whether non-local native passengers and foreign passengers were taken on longer detours than local passengers. Moreover, the exact recording of the route through the GPS-logger allowed calculating the correct price for a given route by applying the tariff system (consisting of distance- and time-related components). By comparing this calculated correct prize with the actually demanded price for the ride, we could measure the frequency and extent of overcharging through demanding more than would be appropriate according to the tariff system. This allowed us then to measure the consequences of being perceived as a foreign passenger not aware of (or of less likely aware of) the details of the tariff system.

In total, we had 348 undercover taxi rides, covering 4,400 km of travel in town, with an average trip length of about 13 kilometers, for which taxi drivers spent about 22 minutes.

Figure 1. *Example for overprovision for foreign passenger (red-and-black route in zig-zag), compared to local passenger (green and straight route). Source data* Balafoutas et al. (2013*)*.

As far as we can say, no taxi driver was requested more than once, which is not surprising, given that there are about 14,000 taxi drivers on service in Athens.

To make sure that our results were not driven by idiosyncracies of specific routes – in many cities the routes to the airport might be considered as particular because they are disproportionally more often frequented by passengers unfamiliar with the city – we let the undercover taxi passengers take many different routes (16 different routes in total). By letting them depart in triples, we could control for traffic and weather conditions. For the analysis of the data, the shortest route in a triple was used to normalize the other two routes to measure the extent of detours (in percentage points). Likewise, the cheapest route within a triple was used to measure the relative costs of informational asymmetries. The existence of overcharging was assessed by checking whether the actual bill was higher than what it should have been conditional on the chosen route.

Figure 1 illustrates a case of overprovision in the form of taking detours. The light and almost straight route was taken by a taxi driver when the undercover customer acted in the role of a local passenger. The dark, zig-zag route was taken with the foreign passenger requesting the same destination. The absolute difference in the trip length was 4.2 km in this (admittedly extreme) case of overprovision.

Overall, both non-local native passengers and foreign passengers were taken on significantly longer detours than local passengers. Out of 13 km average trip length, these detours accounted for roughly 5% of the distance travelled. As expected and predicted, there were no significant differences between non-local natives and foreigners in the extent and frequency of overprovision. Yet, with respect to overcharging, foreigners were ripped off much more often. Overcharging happened in 23% of cases for foreign passengers, but only in 6% of cases for Greek passengers, a highly significant difference. If overcharging occurred, trips became roughly 25% more expensive for a passenger (accounting for about 3 Euro), a considerable increase that was much more costly than overprovision (with

additional costs below 1 Euro on average). While the higher frequency of overcharging of foreign passengers was expected[6], it is worth noting that the reverse side of these results demonstrates a very large fraction of honest behavior of taxi drivers. Specifically, foreigners were in 77% of cases *not* overcharged, which is an amazing frequency given that taxi rides were always one-shot interactions that bear the strongest incentives for exploiting one's informational advantage. Moreover, since taxi drivers in Athens were the residual claimants in all cases, they would have benefitted 1:1 from overcharging, but did not do it even with foreign passengers in the large majority of cases. Truly, though, this fraction of honest charging was even higher for Greek passengers with about 94% of cases.

In sum, the field experiment on taxi services in Athens provided compelling evidence that expert sellers respond to (perceived) informational advantages over the optimal route or the tariff system in a predictable way. While a large fraction of drivers was honest even with passengers with inferior information, the latter were nevertheless more often taken on detours or cheated upon the bill. For those who take a taxi in an unknown city, our results suggest that hiding one's informational disadvantage (by simply mentioning the requested destination, without revealing one's lack of information) could lead to lower prices. Reducing the perceived informational asymmetry may be even better than hiding one's lack of knowledge. One way of conveying the impression to possess relevant information is to instruct the driver which route and main streets to take. This advice seems fairly easy to implement nowadays due to the technology available in modern smart-phones. From a regulator's perspective, modern technologies could also be used to reduce the informational asymmetries and thus the scope for fraud by requiring taxi drivers – or more generally the sellers of credence goods – to make the information about the service they are providing as transparent as possible. In the taxi market, this means, for instance, to require the installment of navigation systems that show the passenger at each point in time the location and driving direction of the taxi on a map of town. Another regulatory response to informational asymmetries would be to impose fixed fares for routes disproportionally often requested by less informed consumers, for example as the ones from the airport to downtown. And this is indeed what we observe in many European capitals.

2.2 On the role of liability and verifiability and the impact of non-standard preferences

The incentives for fraudulent behavior of sellers raise the question how different informational and institutional restrictions affect the provision behavior of sellers. In Dulleck et al. (2011), probably the first lab experiment on the determinants of fraud in credence goods markets, we focused on the effects of liability and verifiability.[7] Liability rules are present

6 As indicated earlier, we ascribe the difference between Greek and foreign passengers to the inferior information of foreigners about the tariff system. One other factor to explain this difference could be a larger willingness of taxi drivers to rip off foreigners as the consequence of a kind of out-group effect (implying that Greek passengers – being the in-group – are better treated). In our case, the out-group coincides with the less informed group, for which reason it is impossible to identify a potential out-group effect in our setting.

7 Somewhat related to the work on credence goods markets are papers by Huck et al. (2012, 2016b) who study the provision of experience goods. These goods (like wine) have characteristics that are unobservable for the consumer *ex ante*, but the quality is revealed after buying or consuming

in many credence goods markets, meaning that the seller is required to provide a service or quality of the good that is sufficient to meet the buyer's demands. Such a rule, while preventing underprovision does not prevent overprovision or overcharging. For instance, the legal regulations for taxi drivers typically stipulate that a taxi driver has to bring the passenger to the requested destination. As Balafoutas et al. (2013) have found, this does not rule out a substantial degree of overprovision in the form of detours or overcharging through inflated bills. One of the research goals in Dulleck et al. (2011) was therefore to investigate the effects of liability on the efficiency of trade on a credence goods market. A second research goal in this project was to study the impact of verifiability on market outcomes. With verifiability we mean that the buyer of a credence good can judge what type of quality or service he has received when buying from an expert seller.

We examined the role of these two factors (i.e. of verifiability and liability) in a large laboratory experiment with a total of 936 participants. While laboratory experiments often raise the issue of external validity of results, they are a means of testing under controlled conditions the effects of single factors in order to isolate the causal effects that determine human behavior under a given set of incentives (Friedman and Sunder, 1994).

In our experiment, we assigned subjects either to the role of a buyer or a seller of a credence good. At the beginning of an interaction period each seller was asked to post the prices for the two (exogenously given) qualities of the good, a high quality and a low one. Given these prices, each buyer could decide whether to interact with the seller or not. If the buyer refused interaction, both the buyer and the seller received an outside option. If the buyer opted for interaction, the seller was informed which type of quality was needed by the buyer to satisfy his or her demands. The seller then delivered a quality and charged one of the two posted prices for it. In the baseline without any institutional restrictions, the seller was not forced to provide a sufficient quality (over- and underprovision were possible), and she was also not forced to charge the price posted for the provided quality (that is, overcharging was also allowed). When verifiability applied, the seller was restricted to charge for the quality that was provided, meaning that overcharging was ruled out. If liability applied, the seller had to provide a sufficient (or even better) quality, meaning that underprovision was prohibited, while overcharging and overprovision was still possible. This stage game was played for 16 periods, with random re-matching of pairs of one buyer and one seller after each period.

The parameters of the market were chosen such that both verifiability and liability induce the buyer to trade with the seller in the theoretical benchmark, thus generating full efficiency and the maximal volume of trade. When neither verifiability nor liability applied, the market was predicted to break down because buyers can expect to fare better with the outside option than when trading with the seller.

Figure 2 shows that the predictions are only partially borne out. In particular, roughly in line with the theoretical prediction, the relative frequency of trade was very high when liability applied. Across all periods, it accounted for 82% of possible trades. Already in the first period there was a large gap between the condition with liability (90% of trade) and the other two conditions included in Figure 2. By contrast, with verifiability, trade occurred in only 63% of cases in the first period, and dropped to an average of 50% across all 16

them and therefore consumers can judge *ex post* whether they received the quality that yields the highest gains from trade or not. The latter is not possible (unless underprovision occurs) with credence goods even after purchase.

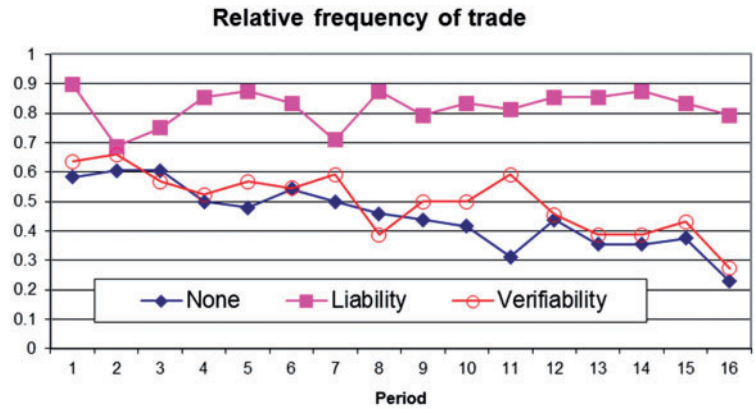

Figure 2. *Frequency of trade when liability or verifiability apply, compared to a situation where none applies. Source data from* Dulleck et al. (2011*)*.

periods. The latter was significantly smaller than when liability applied, although theoretically both conditions should have yielded the same relative frequency of trade. The trade frequency when neither liability nor verifiability applied ("None" in Figure 2) was also contrary to the prediction of complete market breakdown. Rather, it was at 45% across all periods, and as such statistically indistinguishable from the condition with verifiability. Efficiency rates showed a similar picture as the relative frequency of trade. Defining efficiency as the fraction of the maximum total surplus for buyers and sellers that was actually generated, it was 84% with liability, but only 16% with verifiability, and 18% in case of none. Hence, verifiability worked much poorer than predicted, while having none of the two conditions worked much better.

Since these results were puzzling to us at first sight, we examined the underlying causes in more detail in a follow-up paper (Kerschbamer et al., 2015). There, we started from the observation that verifiability had comparatively small effects on efficiency in lab experiments with endogenous price setting of sellers, but the condition without verifiability (and without liability) performed much better than predicted. Only the condition with liability yielded results roughly in line with the theoretical prediction. We identified a lack of robustness of institutional design with respect to heterogeneity in social preferences of credence goods sellers as a possible explanation for these findings. By social preferences we mean that experimental credence goods sellers may not only care for their own material payoff, but may consider the payoffs of their customers as well, when making their provision and charging decisions (see Fehr and Schmidt, 1999, Bolton and Ockenfels, 2000, Charness and Rabin, 2002, or Kerschbamer, 2015 for theoretical models and experimental evidence on the impact of social preferences on behavior). Actually, if some sellers value the welfare of their customers positively (for instance, because they are altruists or have a taste for efficiency) while others have a negative attitude towards them (for instance, because they are inequality averse or spiteful) then this can explain both, why markets with verifiability work worse than predicted, and why markets without work better than predicted.

Key to the argument are the following two observations: First, the standard prediction for credence goods markets with verifiability is that sellers post equal-mark-up prices. Such prices guarantee that sellers have no material incentives to provide an inappropriate

quality. Thus, if sellers are egoistic or pro-social then they behave as predicted. However, sellers with anti-social attitudes will misbehave under equal-mark-up prices. In sum, the standard solution for the verifiability case is robust against the presence of sellers with pro-social other-regarding concerns but non-robust against the presence of sellers with anti-social other-regarding concerns. Second, for markets without verifiability (and without liability) exactly the opposite is true. Here the standard prediction is that all sellers will always provide the lowest quality and charge the price for the highest quality. Since this is already a worst case prediction it is robust against the presence of sellers with anti-social other-regarding concerns but non-robust against the presence of sellers with pro-social other-regarding concerns. Together these two observations provide an explanation for both, why credence goods markets with verifiability fail to reach efficient outcomes and why markets without verifiability (and without liability) perform better than predicted by standard theory.

To test the relevance of this explanation we conducted new experiments (with exogenous prices for the provision of credence goods) that allowed for a parsimonious identification of sellers' social preference types. Based on experimental sessions with 128 subjects, we could show that less than a quarter of subjects behaved according to the standard assumption of selfish payoff maximization, while the rest had either a taste for efficiency or was altruistic – explaining the good performance of markets without verifiability and without liability – while others were inequality averse, spiteful or competitive – explaining the poor performance of verifiability. Overall, our results confirmed the existence of heterogeneity in social preferences which is the main building block for our explanation for the fact that markets with verifiability perform considerably worse than predicted while markets without verifiability (and without liability) perform better than predicted. The liability solution, on the other hand, is a robust institution in the sense that it is unaffected by the presence of sellers with heterogeneous social preferences. This might explain why in the experiments by Dulleck et al. (2011) the condition with liability yielded results roughly in line with the theoretical prediction.

The policy conclusion from Kerschbamer et al. (2015) is that what is needed for a well-performing market is not a perfect institution for one type of agent, but rather an institution that is robust against the coexistence of different types of agents. Verifiability is not sufficient if agents have heterogeneous social preferences, while liability is robust against preference heterogeneity. Since liability requires verifiability of treatment success, while verifiability requires only verifiability of the treatment provided, securing verifiability of success –where possible– might solve credence goods problems more effectively.

Designing robust institutions might not always be possible – recall that liability has also its problem in implementation in practice – meaning that selecting the right agents for jobs involving experts' services becomes particularly important. Instead of choosing doctors, mechanics or computer specialists exclusively according to their training, customers or their representatives should worry more about the attitudes of these experts towards their customers.

A possible objection against deriving policy implications from standard laboratory experiments is that students are different from non-students in many respects and that those differences might translate into different behavior in the field, compared to laboratory experiments. In order to judge the external validity of student data, one possible approach – besides running controlled field experiments – is to compare the behavior of students to that of real professionals in the same laboratory environment. We took this approach in

Beck et al. (2014) by asking whether one would reach similar conclusions regarding the impact of informational and institutional constraints on the behavior of expert sellers on credence goods markets when one took professionals from the target field of interest – the market for car repairs in that case – as participants in lab experiments. So we examined the behavior of 96 car mechanics as sellers in the design of Dulleck et al. (2011) and compared their behavior to the data collected from 140 university students in the role of sellers in exactly the same design.

Reassuringly, we found that car mechanics and students reacted qualitatively very similarly to changes in the informational and institutional framework. This means that liability increased efficiency, while verifiability had no noticeable effect on efficiency, compared to a control condition without verifiability and without liability. The only noticeable difference was that car mechanics were more prone to overprovision in their role as sellers. A possible explanation for this finding is that car mechanics have been repeatedly exposed in their professional career to incentives for overprovision – for instance, because in reality diagnosis is typically subject to errors and because in the car repair market underprovision often leads to negative consequences while overprovision typically remains undetected. This could have made them more prone to providing unnecessary services also in the lab. Whatever the reason is for the higher rate of overprovision, the differences to university students became smaller, and ultimately insignificant, in the course of repetition.

2.3 Does competition in the market matter?

Markets for credence goods are often characterized by competition between sellers. It is unclear, however, whether competition is good or bad for consumers in credence goods markets. On the one hand, it seems reasonable to expect that competition will lead to lower prices which might increase the volume of trade on the market. On the other hand, lower prices may also create stronger incentives for sellers to cheat on consumers. Ex-ante, it is unclear which of these two tendencies dominates and so the consequences of competition on market efficiency are ambiguous.

In the previously discussed experimental study on the effects of liability and verifiability by Dulleck et al. (2011) we also analyzed the effects of competition on the provision behavior of sellers on credence goods markets. This was done in the following way. In all sessions we had matching groups of eight subjects, of which four were in the role of sellers and four in the role of buyers. In half of the sessions we had a bilateral matching, meaning that at the start of each interaction period each seller was randomly and anonymously matched with exactly one buyer from the same matching group. In the other half of the sessions we introduced competition by letting buyers choose their most preferred seller to interact with after the four sellers had posted their prices for the low and the high quality. This means that under competition it could happen that some sellers attracted more than one buyer, while others found no buyer (and thus only earned the value of the outside option). As before, this stage game was repeated for 16 periods. Figure 3 compares – for the setting where neither liability nor verifiability applied (which corresponds to condition "None" in Figure 2 above) – the bilateral matching outcomes to those under competition. Specifically, the figure displays the relative frequency of trade in panel A, the average price for the high-quality good in panel B, and the relative frequency of underprovision in panel C. Panel A shows that with competition, there was significantly more trade between sellers and buyers (73%

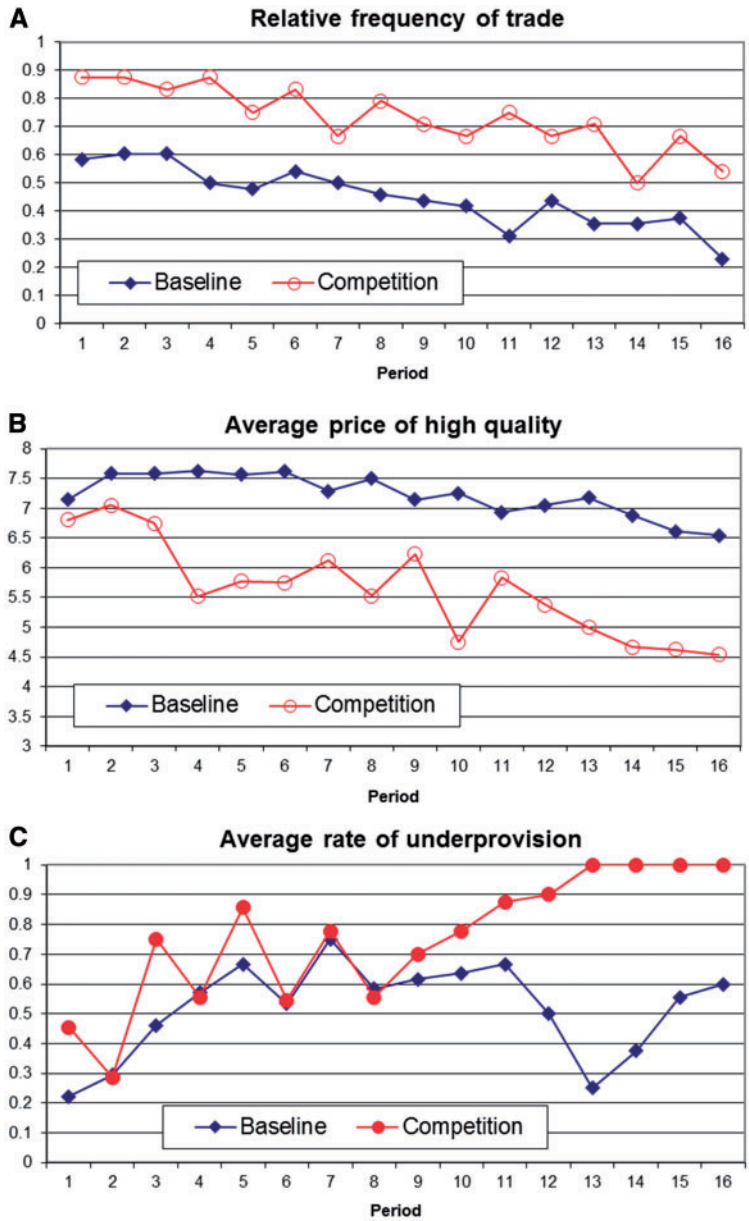

Figure 3. *The impact of competition (among four sellers) on sellers' behavior, compared to bilateral matching of buyers and sellers in Baseline. Source data from* Dulleck et al. (2011)*.*

on average with competition, but only 45% on average with the bilateral matching). This is mainly due to much lower prices under competition. This is shown in Panel B of Figure 3 which confirms that the prices for the high-quality good (which are the empirically relevant prices for markets without verifiability) were much lower with competition (5.7 on average) than with bilateral matching (7.3). Yet, these two beneficial aspects for consumers

were counterbalanced by more cheating with competition. While overprovision and overcharging were not significantly more likely with competition, underprovision happened significantly more often with seller competition than without (73% vs. 53%). The higher frequency of inappropriate provision offset the potentially efficiency-increasing effect of more frequent trade, thus keeping the level of efficiency under both conditions – seller competition vs. bilateral matching – roughly constant.

Mimra et al. (2016) investigated the effects of competition from a different angle than we had done in Dulleck et al. (2011). Instead of comparing market outcomes with exogenous matching to those with endogenous matching, Mimra et al. (2016) compared the provision behavior of sellers when the prices they could charge were determined exogenously and when they could compete in prices by setting them endogenously. Based on an experiment with 320 subjects, the authors found that the level of fraud was higher under price competition than under fixed prices. In particular, the relative frequencies of underprovision and overcharging were significantly higher under price competition than under fixed prices. In addition to the effects of price competition, the authors also studied how the information present on the market – whether a buyer can only identify a seller's behavior towards himself (private history), or whether he can also see how the seller treated other buyers (public history) – affects the sellers' provision behavior. We discuss this aspect of the study in subsection 2.4 (on the effect of reputation).

Rasch and Waibel (2015), both also co-authors in Mimra et al. (2016), used data collected by the German Automobile Association in a field study on car repairs in Germany, and complemented these data with several important key figures for the 134 car garages involved in this study, among them two different dimensions of competition in the market. The first dimension measured the number of competitors in a diameter of 10 kilometers. For this case the authors applied a median split to classify the garages into those with high competition and those with low competition. The second dimension referred to the financial report of the company, particularly whether it had positive or negative equity. This dimension is an indirect measure of competitive pressure in the market. Fraud by garages was captured in Rasch and Waibel (2015) as an indicator variable indicating whether a garage charged for a repair it did not perform – which corresponds to overcharging in the language introduced earlier. Presumably, this variable measured a lower bound for garages' overcharging behavior because it cannot account for the size of overcharging (which could be more or less extensive).

Interestingly, the relative frequency of overcharging was (with 4.5%) relatively low but both measures of competition showed a strong and significant effect on the likelihood of overcharging, yet in opposite directions. The higher the number of competitors in a diameter of 10 kilometers, the lower was the frequency of overcharging. However, a critical financial situation (presumably caused by intense competition in the market) increased the frequency of overcharging significantly, leaving it a bit unclear what is the overall effect of stronger competition on the extent of fraudulent behavior of sellers.

2.4 Is reputation-building sufficient to moderate the effects of asymmetric information?

If sellers and buyers meet repeatedly, as in many real-life credence goods markets, reputational concerns can have a crucial influence on provision and charging behavior. The

pioneering experimental work on the impact of reputational concerns on the performance of credence goods markets is Schneider (2012). In a field study in the market for auto repair service the author examined whether reputational concerns by mechanics mitigate the problems caused by asymmetric information in markets for credence goods. For this purpose the author employs a matched-pair design in which the undercover researcher presented himself either as a one-time consumer or as a repeat-business customer. To present himself as a one-time customer, the undercover researcher entered the garage stating that he was moving away, having moving boxes placed visibly in the back of the car. As a repeat-business customer, the researcher provided a home address near to the garage, adding the remark that he was seeking a local mechanic for an ongoing business relationship.

In total, Schneider (2012) collected data from 91 undercover garage visits with the car having a prearranged set of defects (loose battery cable, low level of coolant, missing taillight). In each case, the mechanic was asked to thoroughly inspect the vehicle, diagnose its condition, make a repair recommendation, and provide a price estimate. The effects of reputation were estimated as the treatment effect on the number of legitimate defects discovered, the diagnosis fee, the repair recommendation, and the repair price.

Overall, the levels of overprovision and underprovision were frighteningly high – in about 30% of cases the mechanics provided completely unnecessary repairs and in about 80% of cases at least one of the pre-fabricated defects was missed. Reputational concerns did not seem to matter, however, contrary to the initial expectation: motorists presenting themselves as repeat-business customers did not receive different repair recommendations, repair prices, or diagnosis quality, compared to one-time motorists. The only difference between the two reputation roles concerned the diagnosis fee, which was 37.70$ for repeat customers, but 59.75$ for one-time customers, indicating a noticeable effect of reputational concerns (in the expected direction) in this dimension of the seller's provision.

Evidence presented by Rasch and Waibel (2015) seems to confirm the slightly positive impact of reputational concerns on consumer-friendly behavior on markets for credence goods. In their field experiment with the German Automobile Association (discussed in the previous subsection) the authors capture the impact of reputational concerns by including a dummy variable standing for the proximity of the garage to the highway in their regressions. Closer proximity to the highway was interpreted as indicating less impact of reputational concerns because of relatively more one-time customers (which is plausible because cars that have a defect on the highway are typically towed to the garage next to the closest exit). The authors found that overcharging was more frequent for garages less than 1.5 kilometers away from the closest highway-exit than for garages further away, indicating again that one-time customers face higher charges in credence goods markets.

Evidence from laboratory experiments is roughly consistent with the finding of a weak positive effect of reputational concerns on consumer-friendly behavior. In Dulleck et al. (2011) – the study on the impact of liability and verifiability on the performance of credence goods markets – we also had an experimental condition where sellers were identifiable for buyers, allowing the latter to keep track of their experiences with a given seller. This means that sellers could build up a reputation with a particular buyer, even though buyers were not informed about this seller's provision and charging behavior towards other buyers. In the baseline condition where neither liability nor verifiability applied, and where the matching was exogenous, the opportunity of reputation-building reduced the likelihood of overcharging considerably, from 88% without reputation-building to 62% with private histories. Figure 4 shows the time path of the overcharging rate across the 16 periods. The

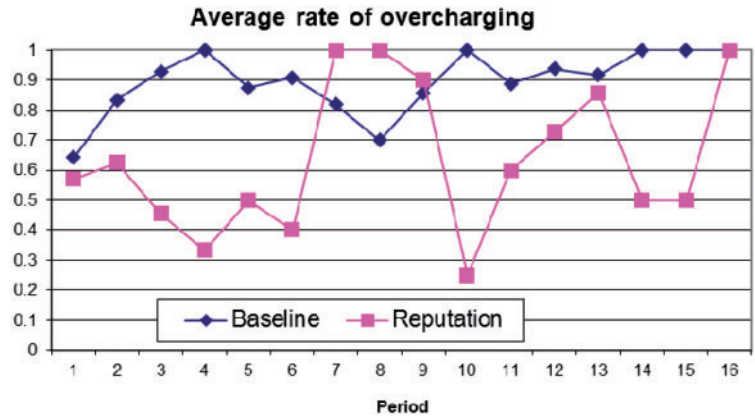

Figure 4. *The impact of reputation-building on sellers' overcharging, compared to Baseline without reputation-building. Source data from* Dulleck et al. (2011)*.*

difference in the frequency of overcharging disappeared, however, as soon as institutional safeguards like liability or verifiability applied.

The experimental study by Mimra et al. (2016) discussed in the previous subsection also investigates the impact of reputational concerns on the provision and charging behavior of experts, but from a slightly different perspective. Specifically, Mimra et al. implemented a 2x2 factorial design varying in one dimension the ability of experts to actively compete for customers via posted prices (this is the treatment variation we have discussed in the previous subsection) and in the other dimension the information of customers regarding experts' success history: In their private histories condition consumers knew for each of their own past interactions which price was charged and whether the repair was successful or not, while in their public histories condition consumers received this kind of information for all consumers and all experts. Comparing the two information conditions, the authors found no significant difference in the underprovison dimension of fraud, while for overtreatment they reported a weakly significant effect whose sign depended on the pricing condition. Here it is important to note that the effect of reputation was captured by comparing a private histories to a public histories condition, while in Dulleck et al. (2011) the impact of reputational concerns was captured by comparing a no histories to a private histories condition.

Overall these results suggest that the effects of an opportunity to build up a reputation as a reliable seller are generally weak and only present in an environment where there are no institutional remedies against fraudulent behavior of sellers. Hence, reputation can at best be considered an imperfect substitute to institutional safeguards if the latter cannot easily be implemented.[8]

8 Theoretical work by Ely and Välimäki (2003) and Ely et al. (2008) suggests that reputational concerns by expert sellers might also lead to less consumer-friendly behavior in markets for credence goods. Grosskopf and Sarin (2010) tested the good and the bad reputation story in a binary trust game experiment with bilateral matching and exogenous cake-sharing rules and found that the positive effects of reputation were generally more widespread than the negative ones. It should be noted, however, that information transmission is very smooth in their artificial framework in which part of the market was played by the computer. It seems therefore difficult to assess how their results translate to a noisier environment.

2.5 How does insurance coverage or third-party reimbursement affect provision?

In many credence goods markets, buyers bear only a part of the costs associated with the provision of credence goods, or are completely sheltered from those costs. Health care insurance is a prime example, and it has been argued for Germany, for instance, that the yearly damage to insurers from faked and inflated medical bills accounted for 6 to 24 billion Euros.[9] Given the importance of this topic and the size of the affected market, it does not come as a surprise that the relationship between health insurance coverage and doctors' prescribing or treatment decisions, or health expenditures more generally, has been a hot topic in the empirical literature for quite some time. Important examples include Lundin (2000), Iizuka (2007 and 2012), Card et al. (2008 and 2009), Wagstaff and Lindelow (2008), Wagstaff et al. (2009), Anderson et al. (2012 and 2014), and Clemens and Gottlieb (2014), and the presented evidence suggests a strong positive correlation between health insurance coverage and medical spending. While these studies nicely exploit exogenous variations of relevant factors to discriminate between several possible explanations for the reported correlation it remains difficult to isolate single factors as causal from non-experimental data. Indeed, the evidence presented in each of these studies is consistent with at least two of the following four hypotheses: (i) individuals with a higher risk of needing medical care are more likely to purchase insurance – an *adverse selection* story; (ii) insured patients ask for more or more expensive service because their out-of-pocket costs are lower – a (first-degree*) moral hazard* story operating at the demand side of the market; (iii) doctors prescribe or provide more or more expensive service to insured patients because they are perfect or imperfect agents for the latter (and because insurance decreases the patient's out-of-pocket costs) – an *agency version of* the standard *moral hazard* story operating at the supply instead of the demand side of the market; and (iv) doctors prescribe or provide more or more expensive service to insured patients because they have material incentives for doing so and expect less resistance from insured agents. Balafoutas et al. (2015) have coined the term "*second-degree moral hazard*" for this latter explanation where a seller in a credence goods market exploits the fact that the relationship between the customer and the third party bearing the costs of the transaction (i.e., the insurance company) is plagued by moral hazard.[10] From a policy point of view, discriminating between these four explanations for the correlation between health insurance coverage and medical spending is of

9 See "Der Spiegel" from 23 December 2012. http://www.spiegel.de/wirtschaft/krankenkassen-detektive-jagen-betruegerische-aerzte-mit-spezialsoftware-a-873059.html (accessed 10 August 2016).

10 We are not aware of any theoretical work investigating this issue. Here it is important to note that the literature on "double moral hazard" addresses a different problem. This literature typically studies settings where two or more agents are engaged in a joint production and where each of them has an incentive to lower the own input. A credence goods example would be a repair where the probability that the product will still work after some time is a function of the diagnosis effort provided by the expert and the maintenance effort provided by the consumer (see Taylor 1995 for such a setting). Note that here the activities of the agents jointly determine the outcome while in the second-degree moral hazard story the activities on the two sides of the market are completely unrelated. Also, while in the double moral hazard literature the moral hazard on one side of the market has a negative effect on the other side of the market, in the second-degree moral hazard story the (expected) moral hazard on the consumer side is exploited by the expert.

primary importance because both the welfare implications and the potential remedies are different.[11]

Evidence that eliminates hypotheses (i) and (ii) as candidate explanations and helps to discriminate between (iii) and (iv) is provided in a recent study by Lu (2014). In a natural field experiment the author randomizes doctors' material incentives (doctors are randomly told either that the patient will buy the prescribed drugs from the doctor or that the drugs will be purchased elsewhere) and patient's insurance status (doctors are randomly told either that the patient has government insurance, or has no insurance). The results indicate that doctors who expect to obtain a fraction of patients' drug expenditures write 43% more expensive prescriptions to insured patients than to uninsured ones, while doctors without a financial incentive do not respond to patients' insurance status.[12] The author interprets this as evidence in support of explanation (iv) – in the language of Balafoutas et al. (2015) this is the second-degree moral hazard explanation – arguing that explanation (iii) would require that doctors without a personal financial incentive also react to patients' insurance status.

Two recent field experiments strongly confirmed the relevance of the second-degree moral hazard mechanism for behavior on markets for credence goods and also showed that in the investigated markets the additional expenditures provided absolutely no benefits for insured customers. In the first one, Balafoutas et al. (2015), we studied whether getting reimbursed from an employer affects the costs of taxi rides. Generally speaking, we studied whether getting reimbursed from an employer affects the costs of taxi rides. Generally speaking, in the relationship between employee and employer, moral hazard kicks in when the employee does not try to minimize costs for the company with regards to expense accounts, travel costs, or company cars, for example, since the employee does not bear the costs of such expenses him- or herself. Actually, the presence of moral hazard in such a business context is an important consideration among practitioners in the field.[13]

In Balafoutas et al. (2015), we collected data on 400 undercover taxi rides in the capital city of Greece, Athens. Our undercover passengers always revealed to the driver that they were unfamiliar with the city in order to create the impression of an informational advantage for the taxi driver. Our main experimental variation consisted solely of a short phrase

11 For instance, in explanations (ii) and (iii) higher medical expenditures for insured patients represent an improvement for patients (although not necessarily an efficient one), while in explanation (iv) higher expenditures are associated with unnecessary or even harmful treatments – implying that the welfare implications are different. Also, in explanations (i) and (ii) the culprit for the correlation between insurance coverage and medical expenditures is the consumer, while in (iii) and (iv) the supply side of the market is responsible for it – implying that the target for policy interventions is different.

12 Similar findings are provided by Iizuka (2007 and 2012). Iizuka (2012) examines micro-panel data from Japan, where some doctors can legally make profits by prescribing and dispensing drugs while others can only prescribe but not dispense them. The author finds that the former doctors respond in their prescriptions to differences in markups while the latter do not and that the former take the patient's out-of-pocket cost into account when choosing between brand-name drugs and generic drugs, while the latter do not. Iizuka (2007) provide similar evidence for hypertension drug expenditures.

13 See, for example, a CNN report on expense account fraud and the extent to which business travelers tend to inflate their expenses (including taxi receipts): http://edition.cnn.com/2011/12/05/travel/expense-account-business-travel/index.html (accessed: 10 August 2016).

added in one of the two experimental conditions. In the treatment condition, the passengers indicated that they needed a receipt in order to have their expenses reimbursed by their employer, while in the control condition the passengers only remarked that they would need a receipt at the end of the ride (without specifying for which purpose).

We found that when reimbursement was mentioned, passengers were about 17% more likely to pay higher-than-justified prices for a given ride. Interestingly, the rate of overprovision (by taking time-consuming detours) did not differ across treatments. Hence, mentioning the reimbursement did not increase overprovision, but had a noticeable effect on the likelihood and the extent of overcharging. This result is very much in line with the earlier findings of Lu (2014) for the health care sector. Here note that getting reimbursed from an employer is somewhat equivalent – in terms of monetary consequences for the consumer – to being insured against the costs of getting some service or buying some product. The second degree moral hazard hypothesis would therefore predict that passengers conveying the impression of not personally incurring the costs of the fare, are more frequently overcharged because taxi drivers expect them to be less likely to notice, or report, fraudulent behavior.

One might argue that the market for taxi rides is special in several respects, for instance in being a highly regulated market (at least in most Western countries), implying that the scope of fraudulent behavior is typically limited. Also, the different sources for fraudulent behavior (overprovision and overcharging) have different consequences for customers in this market than in other important credence goods markets with overprovision putting a burden on the time spent in the taxi, and overcharging having only financial consequences. It seems therefore important to investigate the empirical relevance of second degree moral hazard also in less idiosyncratic markets. This is what we did in Kerschbamer et al. (2016). There we investigated how insurance coverage affected the provision and charging behavior of computer repair shops. For this purpose, we bought identical completely refurbished computers, manipulated one of the two RAM-modules, such that the computer could no longer be booted, and then brought the computers to repair shops. Manipulating the RAM-module has the advantage that it creates an unambiguous error message on the screen, so that every expert working in this field should be able to diagnose the error with certainty, thus ruling out potential incompetence as a possible source for treatment differences.

At the end of the conversation in the shop, the experimenter mentioned in the Baseline condition that he needed a bill – an innocent way of saying that the transaction shall be official, including taxes – while in the Insurance condition he added the information that he needed the bill for an insurance. In total, our undercover clients visited 61 shops, accounting for roughly 25% of the computer repair shops in Austria. Treatment conditions were randomly assigned to different shops. The results were striking to us with respect to the magnitude of the effect that insurance coverage has. We did expect larger bills when mentioning insurance coverage, yet the increase from 70 Euro on average in the Baseline condition to 129 Euro in the Insurance condition was surprisingly large. Figure 5 displays the frequency of repair prices in certain price ranges, showing a strong upward shift with insurance coverage.[14]

14 Note that three of our 61 shops claimed that the computer could not be repaired – a worrying statement given that our RAM-manipulation caused a specific error message. Hence, the number of observations only adds up to 58 in Figure 5.

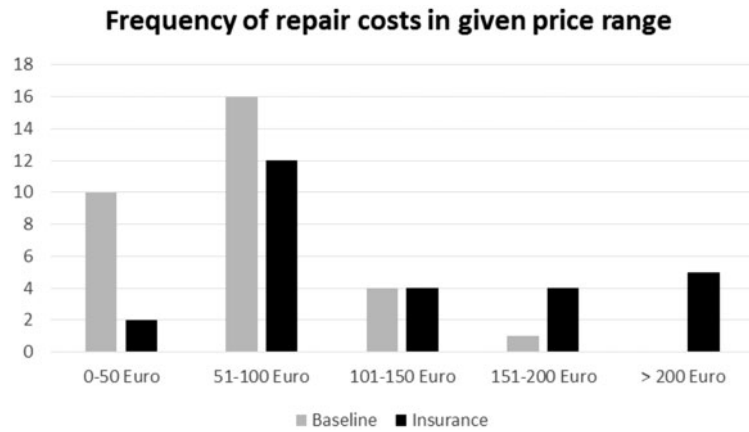

Figure 5. *The impact of insurance coverage on computer repair prices, compared to Baseline without insurance coverage. Source data from* Kerschbamer et al. (2016)*.*

After the repair, we brought the computers to our university's IT-department to let them inspect what had been done in the repair. From this, and from the items listed on the bills, we were able to disentangle the causes of the 80% increase in repair prices in more detail. About one third of the effect was due to overprovision. In fact, overprovision happened *only* in the Insurance condition. If it occurred, repair prices increased quickly to more than 200 Euros. The large bulk of the price increase, namely about two thirds, or 40 Euros, was due to overcharging in working time. Shops in the Insurance condition were much more likely to charge longer working time, on average by about 30 minutes. Since most shops charge by the half hour, and since the repair of a RAM-defect takes about 10-15 minutes only, it is unlikely that the longer working hours charged in the Insurance-condition were caused by mechanics testing extensively (beyond the next 30 minute interval) whether they could find other defects besides the RAM-defect. In sum, the evidence in Kerschbamer et al. (2016) fits nicely into what had been found by Lu (2014) and Balafoutas et al. (2015), showing that if third parties pay the bill from a transaction in a credence goods market, expert sellers tend to exploit this opportunity by inflating the bill.

The evidence from field experiments receives also support from a recent laboratory experiment by Huck et al. (2016a). The authors studied how insurance coverage affects a seller's provision behavior of credence goods and the demand of potential customers. They framed their experiment with 224 subjects as a physician-patient interaction and modeled insurance as an experimental condition in which the potential costs of overprovision by the physician were borne by all patients in a session collectively and not only by the affected patient. The authors found that insurance led to more overprovision, because physicians anticipated that patients were less concerned about the costs of their treatment. In fact, moral hazard loomed on both sides of the market: clients consulted more often (classical moral hazard problem) and expert sellers overtreated more often (second-degree moral hazard) than in a control condition. Adding competition among sellers to insurance coverage only partially offset the adverse effects of insurance, such that most clients still sought treatment, but overprovision of sellers was moderated, with the latter finding being consistent with the field evidence presented by Rasch and Waibel (2015) about the effects of competition in car repair services.

3. Conclusion

Markets for credence goods occupy an important place in any economy. While the goods traded on these markets may be fairly diverse – including health care services, legal or financial advice, repair services, computer programming or simple taxi rides – they share the common feature that expert sellers know more about what the buyer really needs in terms of quality of service or product quality than the buyer him- or herself, and that this holds true not only ex ante, but often also ex post. The informational asymmetry between an expert seller and a buyer creates strong incentives for misbehavior on the sellers' side, at least from a theoretical perspective. Over the past few years, an ever growing body of work has examined under carefully controlled experimental conditions whether these incentives translate into actual misbehavior in the lab and in the field. Summing up the evidence reviewed in this paper, we conclude that expert sellers react in a systematic, and predictable, way to being better informed than their buyers, albeit non-standard preferences often seem to put a limit on the extent of exploiting the informational asymmetries. Institutional safeguards, like liability or verifiability, have been shown to mitigate the problems of overprovision or overcharging, but verifiability has been found to be by far the less effective instrument of the two. Competition on the seller side seems to be a two-edged sword. Judged from the buyer's perspective, it can be both beneficial due to the lower prices it brings, but also costly due to the increased frequency of misbehavior. The ability of sellers to build up reputation works in the expected direction – by reducing the frequency or extent of fraud – but the effects are by far less pronounced than expected and only significant in specific environments. A final aspect covered in our overview – insurance coverage of buyers or, more generally speaking, getting reimbursed by third parties – has shown fairly large, and negative, effects on the provision and charging behavior of expert sellers. In particular, the frequency or extent of overcharging seems to increase considerably in the case of insurance coverage.

Although several important issues have been addressed in recent experimental studies in the lab and in the field, much more work is needed before firm policy conclusions can be drawn from the experimental evidence To start with the example of the effects of insurance coverage, we see two important open questions for future research. First, it is unclear how different insurance schemes affect the provision behavior of sellers and also the behavior of buyers. Partial insurance, respectively co-payment requirements, might have different effects than full insurance. So far, we don't know any systematic study on this aspect. The second important avenue that we see is to understand how insurance companies might effectively cope with the consequences of second-degree moral hazard. One possible approach would be to introduce clauses in insurance contracts that certain services – like repair services – are only reimbursed if provided by repair shops that have a contract with the insurance company, a provision often found in reality. However, we are not aware of any study trying to assess whether such shops will actually cheat less on buyers – and thus on the insurance company – than shops without such contracts. It could be easily the case that the long-term relationship between contracted shops and insurance companies is not sufficient to prevent fraud, which ultimately has to be paid by the buyers through higher insurance premiums.

Another very important question – according to our opinion – concerns the effects of uncertainty in the expert's diagnosis. The laboratory experiments reviewed in this paper were all characterized by the fact that expert sellers could be expected to diagnose the buyer's needs with certainty. This is obviously a harsh assumption that is violated to

different degrees in most real world credence goods markets. The most prominent example for this claim is most likely the health care market where the diagnosis of a patient's needs is very often afflicted by fairly large degrees of uncertainty. Since uncertainty about a patient's needs makes it very difficult to identify the sources and the extent of fraudulent behavior on the seller's side, it remains difficult to judge whether and to which degree the exploding health care costs are driven by overprovision and overcharging or are due to other causes. Most likely, rising costs are caused by a variety of factors, but it would be useful to design field experiments that allow measuring overprovision and overcharging also when experts face uncertainty in their diagnosis of a patient's needs. This would help designing potential countermeasures to limit the degree of fraudulent behavior and thus curtail the increase in health care expenditures.

The provision of credence goods is certainly affected by the extent to which expert sellers profit personally from fraudulent behavior, but too little is known about the exact relationship between incentives and provision behavior. For instance, in most taxi markets in industrialized countries, taxi drivers are the residual claimants of the revenues generated from passengers. This is much less often the case in repair shops – unless they are single-person shops – and comparatively rare in many hospitals where physicians work as employees. In markets for financial advice or legal services, sellers of credence goods are typically compensated in many different formats, including flat payments service, but also large bonus payments in case of a successful acquisition of clients. We think that one future avenue for research will address how different incentive schemes affect the level of appropriate and honest provision of credence goods.

Answers to all of these research questions mentioned in this section will help to design policy interventions with the aim of making credence goods markets more efficient. Certainly, such interventions will have a strong eye on how to protect customers, given their informational disadvantage. Yet, regulation might also overshoot the target of still keeping incentives that make it attractive for expert sellers to offer their services in these markets. To hold the balance between consumer protection and keeping incentives for expert sellers will remain a challenge, but doing more research on the provision of credence goods and therefore gaining a deeper understanding of what drives behavior in these markets will make it easier to meet this challenge successfully.

References

Akerlof, G. A. (1970), "The market for "lemons": Quality uncertainty and the market mechanism", *Quarterly Journal of Economics*. 84, 488–500.

Anderson, M., C. Dobkin, and T. Gross. (2012), "The effect of health insurance coverage on the use of medical services", *American Economic Journal: Economic Policy*. 4, 1–27.

Anderson, M., C. Dobkin, and T. Gross. (2014), "The effect of health insurance on emergency department visits: evidence from an age-based eligibility threshold", *Review of Economics and Statistics*. 96, 189–95.

Balafoutas, L., A. Beck, R. Kerschbamer, and M. Sutter. (2013), "What drives taxi drivers? A field experiment on fraud in a market for credence goods", *Review of Economic Studies*. 80, 876–91.

Balafoutas, L., R. Kerschbamer, and M. Sutter. (2015), "Second degree moral hazard in a credence goods market", *Economic Journal*. forthcoming.

Beck, A., R. Kerschbamer, J. Qiu, and M. Sutter. (2014), "Car mechanics in the lab – Investigating the behavior of real experts on experimental markets for credence goods", *Journal of Economic Behavior and Organization*. 108, 166–73.

Bolton, G. E., and A. Ockenfels. (2000), "ERC: a theory of equity, reciprocity, and competition", *American Economic Review*. 90, 166–93.

Card, D., C. Dobkin, and N. Maestas. (2008), "The impact of nearly universal insurance coverage on health care utilization: evidence from Medicare", *American Economic Review*. 98, 2242–58.

Card, D., C. Dobkin, and N. Maestas. (2009), "Does Medicare save lives? ", *Quarterly Journal of Economics*. 124, 597–636.

Charness, G., and M. Rabin. (2002), "Understanding social preferences with simple tests", *Quarterly Journal of Economics*. 117, 817–69.

Clemens, J., and J. Gottlieb. (2014), "Do physicians' financial incentives affect medical treatment and patient health? ", *American Economic Review*. 104, 1320–49.

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

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

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

Ely, J., D. Fudenberg, and D. Levine. (2008), "When is reputation bad?", *Games and Economic Behavior*. 63, 498–526.

Ely, J., and J. Välimäki. (2003), "Bad reputation", *Quarterly Journal of Economics*. 118, 785–814.

Fehr, E., and K. M. Schmidt. (1999), "A theory of fairness, competition, and cooperation", *Quarterly Journal of Economics*. 114, 817–68.

Friedman, D., and S. Sunder. (1994), *Experimental Methods: A Primer for Economists*. Cambridge University Press.

Grosskopf, B., and R. Sarin. (2010), "Is reputation good or bad? An experiment", *American Economic Review*. 100, 2187–204.

Gruber, J., J. Kim, and D. Mayzlin. (1999), "Physician fees and procedure intensity: The case of Cesarean delivery", *Journal of Health Economics*. 18, 473–90.

Gruber, J., and M. Owings. (1996), "Physician financial incentives and Cesarean section delivery", *RAND Journal of Economics*. 27, 99–123.

Hubbard, T. (1998), "An empirical examination of moral hazard in the vehicle inspection market", *RAND Journal of Economics*. 29, 406–26.

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

Huck, S., G. Luenser, and J. R. Tyran. (2016b), "Price competition and reputation in markets for experience goods. An experimental study", *RAND Journal of Economics*. 47, 99–117.

Huck, S., G. Lünser, and J.R. Tyran (2012), "Competition fosters trust", *Games and Economic Behavior*. 76, 195–209.

Iizuka, T. (2007), "Experts' agency problems: Evidence from the prescription drug market in Japan", *RAND Journal of Economics*. 38, 844–62.

Iizuka, T. (2012), "Physician agency and adoption of generic pharmaceuticals", *American Economic Review*. 102, 2826–58.

Inderst, R., and M. Ottaviani. (2012), "Financial advice", *Journal of Economic Literature*. 50, 494–512.

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

Kerschbamer, R., D. Neururer, and M. Sutter. (2016), "Insurance coverage of customers induces dishonesty of sellers in markets for credence goods", *PNAS – Proceedings of the National Academy of Sciences*. 113, 7454–8.

Kerschbamer, R., M. Sutter, and U. Dulleck. (2015), "How social preferences shape incentives in (experimental) markets for credence goods", *Economic Journal*. forthcoming.

Lu, F. (2014), "Insurance coverage and agency problems in doctor prescriptions: Evidence from a field experiment in China", *Journal of Development Economics*. 106, 156–67.

Lundin, D. (2000), "Moral hazard in physician prescription behavior", *Journal of Health Economics*. 19, 632–62.

Mimra, W., A. Rasch, and C. Waibel. (2016), "Price competition and reputation in credence goods markets: Experimental evidence", *Games and Economic Behavior*. 100, 337–52.

Rasch, A., and C. Waibel. (2015), What drives fraud in a credence goods market? Evidence from a field study. Düsseldorf Institute for Competition Economics Discussion Paper 180.

Schneider, H. S. (2012), "Agency problems and reputation in expert services: Evidence from auto repair", *Journal of Industrial Economics*. 60, 406–33.

Sülzle, K., and A. Wambach. (2005), "Insurance in a market for credence goods", *Journal of Risk and Insurance*. 72, 159–76.

Taylor, C. (1995), "The economics of breakdowns, checkups, and cures", *Journal of Political Economy*. 103, 53–74.

Wagstaff, A., and M. Lindelow. (2008), "Can insurance increase financial risk? The curious case of health insurance in China", *Journal of Health Economics*. 27, 990–1005.

Wagstaff, A., M. Lindelow, G. Jun, X. Ling, and Q. Juncheng. (2009), "Extending health insurance to the rural population: an impact evaluation of China's new cooperative medical scheme", *Journal of Health Economics*. 28, 1–19.

Wolinsky, A. (1993), "Competition in a market for informed experts' services", *RAND Journal of Economics*. 24, 380–98.

Zitzewitz, E. (2012), "Forensic economics", *Journal of Economic Literature*. 50, 731–69.