European University Institute

DEPARTMENT OF ECONOMICS

EUI WORKING PAPERS

ECO 2014/02
Department of Economics

CAR MECHANICS IN THE LAB: INVESTIGATING THE BEHAVIOR OF REAL EXPERTS ON EXPERIMENTAL MARKETS FOR CREDENCE GOODS

Adrian Beck, Rudolf Kerschbamer, Jianying Qiu and Matthias Sutter

European University Institute
Department of Economics

Car Mechanics in the Lab: Investigating the Behavior of Real Experts on Experimental Markets for Credence Goods

Adrian Beck, Rudolf Kerschbamer, Jianying Qiu and Matthias Sutter

EUI Working Paper **ECO** 2014/02

ISSN 1725-6704

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu
cadmus.eui.eu

Car Mechanics in the Lab

Investigating the Behavior of Real Experts on Experimental Markets for Credence Goods*

Adrian Beck[a]

Rudolf Kerschbamer[a]

Jianying Qiu[b,#]

Matthias Sutter[c,d]

[a] *University of Innsbruck*
[b] *Radboud University of Nijmegen*
[c] *European University Institute Florence*
[d] *University of Cologne*

Abstract

Credence goods, such as car repairs or medical services, are characterized by severe informational asymmetries between sellers and consumers, leading to fraud in the form of provision of insufficient service (undertreatment), provision of unnecessary service (overtreatment) and charging too much for a given service (overcharging). Recent experimental research involving a standard (student) subject pool has examined the influence of informational and market conditions on the type and level of fraud. We investigate whether professional car mechanics – as real sellers of credence goods – react in the same way to changes in informational and institutional constraints. While we find qualitatively similar effects in the fraud dimensions of undertreatment and overcharging for both subject pools, car mechanics are significantly more prone to supplying unnecessary services in all conditions, which could be a result of decision heuristics they learned in their professional training.

JEL classifications: C91, D82, C72

Keywords: artefactual field experiment, car mechanics, credence goods

This Version: December 25th, 2013

---

* We thank Dirk Engelmann as the co-editor in charge, an anonymous associate editor and two anonymous referees for their constructive reports. Their thoughtful comments and suggestions helped to improve the paper a lot. Financial support from the Austrian Science Fund (FWF) through grant number P20796 and from the Australian Research Council (ARC) through grant number DP110103653 is gratefully acknowledged.

Corresponding author: Thomas van Aquinostraat 5.1.12, 6525 GD Nijmegen, The Netherlands, Email: j.qiu@fm.ru.nl

1. Introduction

Credence goods are characterized by asymmetric information between sellers and consumers on the quality of service that yields the highest surplus from trade: While sellers learn that quality by performing a diagnosis, consumers are unable to judge which quality is the surplus maximizing one. Moreover, in many cases consumers are not even *ex post* able to observe the received quality. The seminal paper on credence goods is by Darby and Karni (1973), who added this type of good to Nelson (1970)'s classification of ordinary, search and experience goods. Typical examples of credence goods are car repairs, medical services, software programming, or taxi rides in an unfamiliar city. Hence, despite of the uncommon name, credence goods are frequently consumed and economically important.

The informational asymmetries prevalent in markets for credence goods invite fraudulent behavior by sellers, implying that the search for institutions that increase efficiency on credence goods markets is a highly relevant topic in economics (see Dulleck and Kerschbamer, 2006, for a survey of the literature). In particular, credence goods markets typically suffer from the following types of cheating on consumers: (1) *undertreatment,* i.e., providing a quality that is insufficient to satisfy the consumer's needs; (2) *overtreatment,* i.e., choosing a higher quality than the surplus maximizing one; and (3) *overcharging,* i.e., charging for a higher quality than has been provided.

That fraud is more than mere a theoretical possibility in markets for credence goods has been documented, among others, by Domenighetti et al. (1993). The authors examine how patients' information affects the treatment they receive in hospitals and find that common surgical procedures are less frequent for patients perceived as being better informed. In a similar vein, Gruber and Owings (1996) show that the relative frequency of cesarean deliveries responds to the remuneration for it. While these and other field studies impressively document the existence of fraud, a general disadvantage of field data is the lack of controlled variation of factors predicted to be crucial by theory. Controlled variation is the key advantage of the laboratory. In the context of credence goods lab experiments have recently been conducted by Kerschbamer et al. (2009) and Dulleck et al. (2011). Those studies examine the influence of informational and market conditions on the type and extent of fraud, finding, among others, that liability clauses (preventing undertreatment) are key for the efficient provision of credence goods, whereas verifiability (preventing overcharging) fails to improve efficiency, although in theory it should. While providing important information on the impact of institutions on market outcomes, the experimental studies by Kerschbamer et al. (2009) and Dulleck et al. (2011) leave one important question unanswered: Do real world sellers of credence goods react in the same way to changes in the informational and institutional framework as university students do? This question touches on the issue of external validity of laboratory data.

In principle, there are two ways to address this question. The first one is to study the behavior of professionals in field experiments. Schneider (2012) and Balafoutas et al. (2013) are examples for this approach. Schneider (2012) brought his car for repair to different garages, sometimes suggesting the potential for repeated interaction in the future, sometimes inducing the impression that repeated interaction was highly unlikely. Based on data from 91 undercover garage visits, the author finds no evidence that a mechanic's concerns for reputation have an influence on the service provided; however it has an impact on diagnosis fees. Balafoutas et al. (2013) have studied the impact of perceived information on the type and extend of fraudulent behavior of taxi drivers. Based on the data from more than 300 undercover taxi rides, the authors find that taxi drivers exploit their informational advantage in a systematic way by taking passengers perceived as uninformed about the city on longer detours and charging unjustified surcharges to passengers perceived as uninformed about the tariff system.

While these field studies provide compelling evidence about the problems prevalent in credence goods markets, they do not directly address the question of external validity of results based on lab data generated with a university student subject pool. In particular, it is often argued that students are different from non-students in many respects and that those differences might translate to different behavior. So, for judging the external validity of student data one way is to compare the behavior of students to that of real professionals in the same environment. This is the way the issue of external validity is addressed in the current paper.[1] Specifically, we ask the question whether one would reach similar conclusions regarding the impact of informational and institutional constraints on the behavior on markets for credence goods by taking professionals from the target field of interest –the market for car repairs in the present case– as participants in lab experiments. In addition, we are also interested in quantitative differences across subject pools for a given institutional framework, but only to the extent that those differences have economically relevant implications for optimal institutional design. Addressing those issues seems important because the ultimate goal of lab experiments in the context of credence goods is to complement theoretical work in search for institutions that help to contain the amount of fraud in real world credence goods markets and because in the end experts –and not students– make the key decisions in such markets.

[1] A related external validity issue is that the lab is an artificial environment and as such might miss some behaviorally relevant features of the field. This dimension of external validity is beyond the scope of this paper.

To address those issues we let 96 car mechanics take decisions as sellers in an experimental credence goods game and compare their behavior to that of 140 university students in the role of sellers. We find that car mechanics react qualitatively very similar to changes in the informational and institutional framework as students do. Regarding quantitative differences across subjective pools within a given informational and institutional framework our most important finding is that car mechanics have a more pronounced tendency to supply unnecessary services in each institutional framework. We argue that this difference in behavior is probably due to decision heuristics car mechanics learned in their professional training.

Several studies have compared the behavior of professionals and students in other environments. Many of them find that professionals' behavior is qualitatively similar to that of students. Examples include Siegel and Harnett (1964) who compare the behavior of students and employees in the industrial sale operation division of General Electric in a bargaining game and find that the two subject pools behave largely similar; Dyer et al. (1989) who compare the behavior of students to that of executives from construction companies in common value auctions and find that both exhibit the winner's curse and share also other relevant patterns; Cooper et al. (1999), who compare the behavior of students and managers in a market entry game, finding similar core behavior; and Potters and van Winden (2000) who compare the behavior of students and public relationship officers in a lobbying game and only find minor difference.[2] Those studies finding differences in behavior across professionals and students largely attribute them to decision rules, abilities, preferences, or information that professionals have acquired during their career. Burns (1985), for instance, attributes the difference in behavior between experienced wool buyers and students in a progressive oral auction to the decision rules the former have acquired in the market they know; Fehr and List (2004) report CEOs to be more trusting and more trustworthy in an experimental trust game than students are, arguing that CEOs have experienced more often the possible efficiency gains from trust and trustworthiness; Alevy et al. (2007) attribute the difference in behavior between market professionals from the floor of the Chicago Board of Trade and students in an experimental information cascade game to the fact that the former use a more sophisticated decision process and are better able to discern the quality of public signals; and Carpenter and

[2] The impression that the majority of studies comparing the behavior of students and experts in standard lab experiments find qualitatively similar patterns is confirmed by Fréchette (2013). The author reviews 13 such studies and concludes (on page 33) that "[i]n 9 of those 13, professionals are not closer or further from the theory in a way that would lead us to draw different conclusions." See Ball and Cech (1996) for a broader survey on subject pool differences in general, and Camerer (2013) for an even broader discussion on external validity of lab experiments. Al-Ubaydli and List (2013) propose a theoretical model that helps frame the crucial features in the debate on lab-field generalizability.

Seki (2011) compare the behavior of shrimp fishermen organized in a group that shares both income and operating expenses to that of students in a voluntary contribution game and attribute the finding that the former contribute significantly more to endogenously acquired social preferences.

The rest of the paper is organized as follows. Section 2 introduces a simple model of a credence goods market. Section 3 presents the experimental procedure and Section 4 the results. Section 5 concludes.

2. The Credence Goods Model

Consider the following credence goods game, henceforth referred to as condition ***B*** (for baseline). It is a simplified and parameterized version of a game studied by Dulleck and Kerschbamer (2006) and it corresponds to condition B/N in Dulleck et al. (2011). There are two players, an expert seller and a customer. First, the expert posts two prices: P^L for a low quality service and P^H for a high quality service, where P^L, $P^H \in \{1, 2, \ldots, 11\}$ and $P^L \leq P^H$. The customer gets to know these prices and then decides whether to stay out of the market or to interact with the expert. If the customer stays out, the game ends and both parties receive an outside option of 1.6 points. If the customer decides to interact with the expert, nature assigns with equal probability the need for the low and the need for the high quality service. The expert learns the customer's need and provides either the low (q^L) or the high (q^H) quality, where the low (or high) quality has costs of $c_L = 2$ ($c_H = 6$, respectively) points. Finally, the expert charges one of the posted prices: P^L if he claims to have provided the low quality, and P^H otherwise. The customer receives a value of 10 points from the interaction if she receives a sufficient quality (i.e., if she needs the low quality and gets either the low or the high quality, or if she needs the high quality and gets it), otherwise she receives a value of zero. Whether or not the customer's need is satisfied, she must pay the price of the quality the expert claims to have provided. Figure 1 illustrates the sequence of moves in this game and the material payoffs of each player.

In real credence goods markets various informational and institutional conditions might limit a seller's fraudulent behavior. We consider two alternatives to our baseline condition, both limiting a seller's strategy space (in comparison to condition ***B)***: (1) *liability* (condition ***L***), requiring that the seller provides a sufficient quality, and (2) *verifiability* (condition ***V***), requiring that the seller charges the price for the quality of service rendered. Thus, liability precludes *undertreatment*, and verifiability precludes *overcharging*. Note that liability does not preclude overcharging, whereas verifiability does not prohibit undertreatment. Neither of them prevents *overtreatment*.

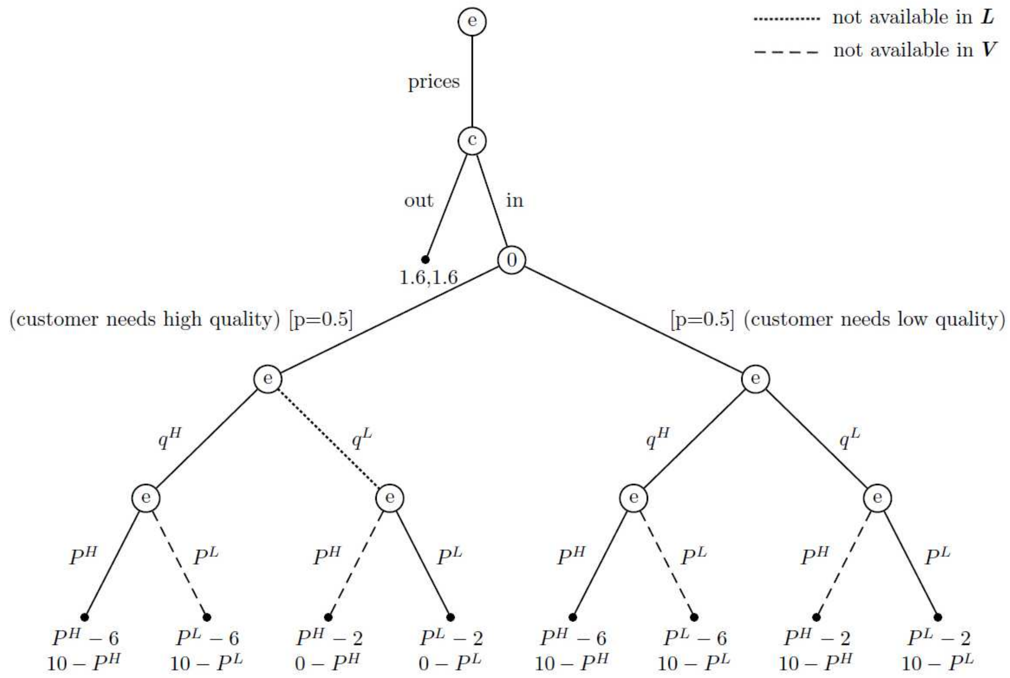

Figure 1: Game Tree

Assuming common knowledge that all players are rational and only interested in own material payoffs the three games can be solved via backward-induction. In condition ***B***, an own-money maximizing expert always provides the low quality and charges for the high one. Anticipating this, the customer only enters if $P^H \leq 3$. But with such a P^H an expert is unable to earn the value of the outside option. Thus, the market breaks down. In condition ***L***, an own-money maximizing expert always provides the appropriate quality and charges P^H. Anticipating this, the consumer accepts if $P^H \leq 8$. Thus, the expert posts $P^H = 8$ (P^L is indeterminate) and the customer enters the market. In condition ***V***, the own-money maximizing expert always chooses (and charges for) the quality with the higher mark-up (defined as the difference between price and cost) and is willing to choose the appropriate quality if markups are the same for both qualities. Since consumers correctly anticipate this, the expert cannot gain by cheating. Thus, the expert posts the most profitable equal-markup price vector that is accepted – which is the vector (P^L, P^H) = (6, 10) – and the customer enters the market. Table 3 in the Appendix summarizes those benchmark predictions.

3. Experimental Procedure

We recruited car mechanics in their third or fourth year of professional training from the “Tiroler Fachberufsschule für Kraftfahrzeugtechnik”, a vocational school located in Innsbruck, Austria. These mechanics are apprentices and work as regular employees at their respective auto repair shops. We use them as experts in our experiment. Students at the University of

Innsbruck acted as customers. The experiment was conducted with our mobile lab at the vocational school. In total, 96 car mechanics acted as sellers and 96 undergraduate students acted as customers in the computerized experiment using zTree (Fischbacher, 2007). At the start of each session, we informed all subjects that there are two roles in the experiment and that car mechanics would exclusively act in one role and students would exclusively act in the other. Then the game was explained to subjects. Given that car mechanics have less education and therefore may have difficulties in understanding the experimental procedure, we took great care to make sure they fully understood the game. We used an intuitive graph to illustrate the sequence of decisions to car mechanics (see Figure 2 in the Appendix). We walked car mechanics through the games slowly and explained to them the implications of each possible choice. Also, all subjects had to answer a set of control questions individually, and the experiment proceeded only after all control questions had been answered correctly. We had 3 conditions in the experiment: the baseline condition (condition ***B***), the liability condition (condition ***L***), and the verifiability condition (condition ***V***). For each condition, we ran 2 sessions with 32 subjects each (16 car mechanics and 16 students). Within a session, we formed independent matching groups of four car mechanics and four students, with random re-matching of one car mechanic and one student after each round within each matching group. The stage game was played 16 rounds, and earnings were accumulated across all rounds. The exchange rate between experimental points and euros was 0.25 euro per point, and participants earned about 13 euros in less than 90 minutes.

As a comparison to these sessions with car mechanics as experimental sellers of credence goods, we use a subset of the data presented in Dulleck et al. (2011). A few months prior to our experiment with car mechanics, they ran an experiment with an almost identical procedure and the same monetary incentives, using a subject pool consisting of 280 university students, 140 of which were in the role of sellers.[3]

4. Experimental Results

Table 1 provides descriptive statistics for all experimental treatments, where ***B***, ***L*** and ***V*** indicate the three experimental conditions introduced above, and subscripts ***C*** and ***S*** refer to

[3] There are three differences in the experimental design between Dulleck et al. (2011) and the present study: (1) In the present study all car mechanics were assigned to the role of experts and all students were assigned to the role of consumers, while in Dulleck et al. roles were assigned randomly. (2) In the present study students received an extra show-up fee of 5 Euros to compensate for the additional traveling time; the variable part of subjects' earnings was exactly the same as in Dulleck et al. (2011). (3) In the present study the sequence of decisions in the stage game was additionally illustrated by displaying the "decision tree" in Figure 2 while in Dulleck et al. (2011) the instructions were purely verbal.

car mechanics and students in the role of sellers.[4] Interaction is calculated as the proportion of cases where customers agree to interact. Efficiency is calculated as the ratio of the actual average profit and the maximum possible average profit per period (adjusted by subtracting the outside option from the nominator and denominator). Undertreatment (overtreatment) is calculated as the relative frequency with which the expert provides the low (high) quality when the customer needs the high (low) one. In order to make overcharging comparable between conditions ***B*** and ***L***, we follow Dulleck et al. (2011) in defining overcharging as satisfying the following conditions: (*i*) $P^H > P^L$, (*ii*) the customer needs the low quality, and (*iii*) the expert provides the low quality, but charges for the high one.

Table 1: Descriptive Statistics

	car mechanics			students[#]		
	$\boldsymbol{B_C}$	$\boldsymbol{L_C}$	$\boldsymbol{V_C}$	$\boldsymbol{B_S}$	$\boldsymbol{L_S}$	$\boldsymbol{V_S}$
P^L	4.79	5.81	5.61**	4.95	5.98	6.03
P^H	7.47	8.22	8.06*	7.63	8.14	7.76
interaction	0.50	0.67***	0.50	0.45	0.82	0.50
efficiency	0.11	0.62***	0.13	0.18	0.84	0.16
undertreatment	0.61	n/a	0.49*	0.53	n/a	0.60
overtreatment	0.20***	0.11**	0.25***	0.06	0.02	0.05
overcharging	0.72	0.67	n/a	0.88	0.75	n/a
number of subjects	64	64	64	96	96	88

*** / ** / * difference between subject pools significant at the 1% / 5% / 10% level according to two-sided Wilcoxon rank-sum test using group averages as the unit of observation
[#] data taken from Dulleck, Kerschbamer and Sutter (2011).

Result 1*. Regardless of whether experts are car mechanics or students, (i) markets under condition* ***B*** *perform better than standard theory predicts, (ii) markets under condition* ***V*** *perform worse than standard theory predicts, and (iii) imposing* ***L*** *has a large and significant effect on interaction rates and efficiency, while imposing* ***V*** *does not.*

Support for R1 (i): The interaction ratio is 0.50 in $\boldsymbol{B_C}$, respectively 0.45 in $\boldsymbol{B_S}$, which is in sharp contrast to the theoretical prediction of no interaction, but is consistent with existing empirical evidence of the functioning of actual car repair markets (c.f. Schneider 2012, for instance). Actual undertreatment and overcharging ratios are significantly below the predicted level of 100%, which increases incentives for customers to enter the market. However, customers are, in general, too credulous as they enter the market although it does not pay to do so in the aggregate. Nevertheless, efficiency levels are clearly above the theoretical prediction of zero with both subject pools. Dulleck et al. (2011) discuss social or moral norms

[4] Throughout the market conditions **B, L** and **V** are called ‘conditions’, while a market condition combined with a subject pool is called a ‘treatment’.

and non-trivial distributional preferences as possible explanations. Kerschbamer et al. (2009) explore the latter hypothesis more systematically and provide supporting evidence for that explanation.

Support for R1 (*ii*) and (*iii*): The benchmark solution (displayed in Table 3 in the Appendix) suggests that both liability and verifiability should increase interaction rates and efficiency up to 100%. However, only liability, but not verifiability, increases the interaction rate and efficiency significantly in comparison to condition ***B***, controlling for the subject pool. More precisely, interaction rates are significantly increased in condition ***L*** ($p < 0.01$ for B_S vs. L_S and $p < 0.05$ for B_C vs. L_C), but not in condition ***V*** ($p > 0.33$ for B_S vs. V_S and $p > 0.91$ for B_C vs. V_C). Efficiency is significantly increased in condition ***L*** ($p < 0.01$ for B_S vs. L_S and $p <$ 0.01for B_C vs. L_C), but not in condition ***V*** ($p > 0.53$ for B_S vs. V_S and $p >$ 0.64for B_C vs. V_C).[5] The insignificant effect of verifiability on interaction – compared to condition ***B*** – is driven by two factors. First, the prediction for ***V*** is based on expert sellers posting equal-markup vectors (i.e., price-vectors with P^H - c_H = P^L - c_L). Such vectors are very rarely chosen in $\boldsymbol{V_C}$ (13%) and in $\boldsymbol{V_S}$ (4%), though. Second, the benchmark solution assumes that experts provide the appropriate quality under equal-markup price-vectors. However, experts frequently over- or undertreat even under such vectors: Students provide an inappropriate quality in 17% of the cases under equal-markup vectors, while car mechanics do so in 42% of the cases.

Comparing the behavior of the two subject pools within a given condition our most remarkable finding is an economically important and statistically significant difference in overtreatment rates in all conditions:

Result 2. *Regardless of the informational and institutional conditions under which transactions take place, car mechanics provide overtreatment much more often than student experts.*

Support for R2: In all conditions, car mechanics provide the high quality when the low one is needed significantly more often than students ($p < 0.05$). Notice that under conditions ***B*** and ***L***, overtreatment causes unnecessary costs for the expert without any compensating benefit. Thus, it is pure waste at an own material cost in these conditions. Only in condition ***V*** overtreatment can be profitable for sellers; specifically it is profitable if they have posted overtreatment price-vectors (i.e., price-vectors with P^H - c_H > P^L - c_L). In treatment $\boldsymbol{V_S}$ we observe such price-vectors in only 2% of all cases, while we observe them in 13% of cases in treatment $\boldsymbol{V_C}$. Thus, in condition ***V*** the higher frequency of overtreatment by car mechanics

[5] All tests are two-sided Wilcoxon rank-sum tests using group averages as the unit of observation. Group here refers to an independent matching group consisting of 4 sellers and 4 consumes.

can (at least partly) be explained by their price-posting behavior. In Table 1 the difference in price-posting behavior in condition ***V*** manifests itself in the price spread between high and low quality, which is larger, on average, for car mechanics. The difference in price-posting behavior between $\boldsymbol{V_C}$ and $\boldsymbol{V_S}$ might also be responsible for another (weakly significant) difference in behavior in condition ***V***: Car mechanics provide the low quality when the high quality is needed in 49% of the relevant cases, while students do so in 60% of the cases. While the difference between subject pools in posted prices can at least partly explain the difference in under- and overtreatment frequencies between subject pools in condition ***V*** it cannot explain the high overtreatment frequency of car mechanics in the other two conditions. Table 1 reveals another difference in the behavior of the two subject pools: The interaction rate is significantly lower in $\boldsymbol{L_C}$ than in $\boldsymbol{L_S}$**.** This is surprising as with liability consumers do not have to worry about undertreatment and because the posted prices are quite similar in the two treatments. A possible explanation for this finding is that the higher overtreatment frequency in $\boldsymbol{L_C}$ as compared to $\boldsymbol{L_S}$ drives consumers out of the market. A direct consequence of the higher overtreatment rate and the lower interaction rate in $\boldsymbol{L_C}$ is that efficiency is lower in $\boldsymbol{L_C}$ than in $\boldsymbol{L_S}$**.**

Returning to overtreatment, we ran a Probit regression with random effects on individual subjects for each condition to examine the phenomenon more carefully. The dependent variable is the probability of overtreatment. As independent variables we use P^L, P^H, period, and a dummy variable indicating whether the expert was a car mechanics. The results of these regressions are reported in Table 2. As expected from the non-parametric tests, the car mechanics dummy is positive and highly significant (at the 1%-level) in each condition, confirming that car mechanics are more prone to provide overtreatment than students. Also, in all conditions overtreatment decreases in P^L and this effect is particularly pronounced in condition ***V*****.** In this latter condition overtreatment also strongly increases in P^H. For condition ***V*** those price effects are perfectly in line with the material incentives the sellers face, while in conditions ***B*** and ***L*** they are only in line with the material incentives if seller intend to charge the price for the quality provided (which some seller do, but not many; this might explain why the price effects are much weaker in conditions ***B*** and ***L*** than in ***V***). Finally, the negative coefficient of the variable *period* indicates that overtreatment decreases over time. This pattern is confirmed by Figure 3 (in the Appendix), which shows the development of the overtreatment ratio over the last 10 periods in the six treatments. Overtreatment ratios of car mechanics start much higher and remain much higher – despite falling over time – than those of student sellers.

Table 2: A Random Effects Probit Regression

dependent variable	overtreatment		
market condition	***B***	***L***	***V***
P^L	-0.144* (0.081)	-0.121* (0.065)	-0.545*** (0.102)
P^H	0.0541 (0.073)	-0.142 (0.100)	0.484*** (0.117)
period	-0.080*** (0.026)	-0.077*** (0.026)	-0.067** (0.028)
car mechanic	0.699*** (0.260)	0.697*** (0.238)	0.912*** (0.284)
constant	-0.837* (0.475)	0.275 (0.749)	-2.085** (0.863)
observations	301	502	299
prob > chi2	0.000	0.000	0.000

*, **, *** denotes significance at the 10%, 5%, 1% level respectively
Note: Random effects are on the level of the individual subject. Interaction terms were dropped due to their insignificance and the improvement in model fit (Bayesian information criterion).

Turning to possible explanations for our results, students differ from car mechanics in many dimensions including family background, wealth, education, etc. and each of these differences could potentially be responsible for the differences in behavior we observe. If they are, then our findings are not primarily due to the fact that one group is composed of professionals in the field of interest while the other is composed of students. However, with most of the mentioned factors it seems unlikely that they directly drive our findings. More likely, car mechanics' professional training, routines, and norms induce them to use decision heuristics different from those of students. In the current experiment, car mechanics are introduced to the new, unfamiliar environment of an economic experiment. When facing such an environment it seems plausible that car mechanics rely on decision heuristics for similar situations experienced in the field. In that sense, the tendency of car mechanics to overtreat seems to be consistent with the incentives they face in everyday life. In the car repairing business undertreatment often implies serious consequences, either due to institutional constraints or the car mechanic's concern for his reputation. Overtreatment, on the other hand, is difficult to detect and to punish. This might yield strong incentives in the car repair business to provide higher quality than warranted. And there is indeed evidence suggesting that these incentives translate in behavior. For instance, the US Department of Transportation estimated that more than half of the expenses for car repairs are unnecessary, thus constituting overtreatment (Wolinsky, 1993). It seems reasonable that our car mechanics have been repeatedly exposed in their professional career to this kind of incentives for overprovision, making them more prone to providing unnecessary services also in our experiment.

5. Conclusion

We have compared the behavior of university students and car mechanics as sellers in an experimental credence goods market. Our results have shown that both subject pools react qualitatively in the same way to changes in the informational and institutional framework. Specifically, we have found that liability – but not verifiability – has a strong effect on the likelihood of interaction and overall market efficiency. Thus, the main results of previous experimental studies (by Kerschbamer et al., 2009, and Dulleck et al., 2011) on the effects of institutional safeguards on credence goods markets are robust against a change in subject pool from university students to real experts. Regarding differences in behavior across subject pools within a given market condition we have found an economically important and statistically significant difference in the propensity to provide too high quality (like unnecessary repairs). Specifically, car mechanics have chosen overtreatment significantly more often than university students in all conditions. In line with previous literature (see, e.g. Burns, 1985; Alevy, et al., 2007), we have argued that this difference is most likely due to particular decision rules that professionals – car mechanics in our case – have developed in the past. Since in the car repairing business – as in many other credence goods markets – undertreatment is typically observable and verifiable while overtreatment is not, and since in real life a diagnosis is typically subject to errors, real experts often face strong incentives for providing “safe solutions”. Seen from that perspective, the tendency of car mechanics to overtreat in an experimental market for credence goods seems to be consistent with the incentives they face in their everyday professional life.

We view our findings as potentially having important implications for the design of institutions for credence goods markets: Previous experimental results (by Kerschbamer et al., 2009; Dulleck et al., 2011, or Beck et al., 2013, for instance) have suggested that overtreatment is hardly an issue in such markets while common sense suggests that it is an important phenomenon in real credence goods markets. If the findings reported here are confirmed in future research then they suggest that imposing liability alone is insufficient to contain fraud in markets for credence goods and this insight might then stipulate further theoretical work on the design of better institutions. Besides providing important substantive information on behavior in markets for credence goods, the present work (together with Dulleck et al. 2011) also offers a methodological contribution in suggesting the complementary role of experiments with students (which can easily be conducted on a large scale) and those with professionals (where the scale of the experiment is typically limited).

Investigating how professionals make decisions in stylized mini-games in the lab touches only one dimension of external validity, of course. A related issue concerns the artificial environment in the lab, which also merits serious consideration. In that respect our work presents only a first step and more work is needed to bridge the gap between the lab and the naturally-occurring setting of interest, before conclusions can be drawn for the optimal design of institutions for markets for credence goods.

References

Alevy, J., Haigh, M., List, J., 2007, Information Cascades: Evidence from a Field Experiment with Financial Market Professionals, *Journal of Finance* 62, 151-180.

Al-Ubaydli, O., List, J.A. (2013) On the Generalizability of Experimental Results in Economics: With A Response To Camerer, *NBER Working Paper* 19666, November 2013.

Balafoutas, L., Beck, A., Kerschbamer, R., Sutter, M., 2013, What Drives Taxi Drivers? A Field Experiment on Fraud in a Market for Credence Goods. *Review of Economic Studies* 80, 876-891.

Ball, S., Cech, P.A., 1996, Subject Pool Choice and Treatment Effects in Economic Laboratory Research, *Research in Experimental Economics* 6, 239-292.

Beck, A., Kerschbamer, R., Qiu, J., Sutter, M., 2013, Shaping Beliefs in Experimental Markets for Expert Services: Guilt Aversion and the Impact of Promises and Money-Burning Options, *Games and Economic Behaviour* 81, 145-164.

Burns, P., 1985, Experience and Decision Making: A Comparison of Students and Businessmen in a Simulated Progressive Auction, *Research in Experimental Economics* 3, 139-153.

Camerer, C. F., 2013, The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List, in G. Fréchette and A. Schotter (eds.) *The Methods of Modern Experimental Economics*, Oxford University Press (in press).

Carpenter, J., Seki, E., 2011, Do Social Preferences Increase Productivity? Field Experimental Evidence from Fishermen in Toyoma Bay, *Economic Inquiry* 49 (2), 612–630.

Cooper, D., Kagel, J., Lo, W., 1999, Gaming against Managers in Incentive Systems: Experimental Results with Chinese Students and Chinese Managers, *American Economic Review* 89, 781-804.

Darby, M. R., Karni, E., 1973, Free Competition and the Optimal Amount of Fraud, *Journal of Law and Economics* 16, 67-88.

Domenighetti, G., Casabianca, A., Gutzwiller, F., Martinoli, S., 1993, Revisiting the Most Informed Consumers of Surgical Services: The Physician-Patient, *International Journal of Technology Assessment in Health Care* 9, 505-513.

Dulleck, U., Kerschbamer, R., 2006, On Doctors, Mechanics, and Computer Specialists: The Economics of Credence Goods. *Journal of Economic Literature* 44, 5-42.

Dulleck, U., Kerschbamer, R., Sutter, M., 2011, The Economics of Credence Goods: On the Role of Liability, Verifiability, Reputation and Competition, *American Economic Review* 101(2): 526–555.

Dyer, D., Kagel, J. H., Levin, D., 1989, A Comparison of Naive and Experienced Bidders in Common Value Offer Auctions: A Laboratory Analysis, *The Economic Journal* 99(394), 108-115.

Fehr, E., List, J., 2004, The Hidden Costs and Returns of Incentives-Trust and Trustworthiness Among CEOs, *Journal of the European Economic Association* 2, 743-771.

Fischbacher, U., 2007, Z-tree: Zürich Toolbox for Readymade Economic Experiments. *Experimental Economics* 10, 171-178.

Fréchette, G. R., 2013, Laboratory Experiments: Professionals versus Students, in G. Fréchette and A. Schotter (eds.) *The Methods of Modern Experimental Economics,* Oxford University Press (in press).

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

Kerschbamer, R., Sutter, M., Dulleck, U., 2009, The Impact of Distributional Preferences on (Experimental) Markets for Expert Services, IZA Discussion Paper No. 4647. IZA Bonn.

Nelson, P., 1970, Information and Consumer Behavior, *Journal of Political Economy* 78, 311-29

Potters, J., van Winden, F., 2000, Professionals and Students in a Lobbying Experiment: Professional Rules of Conduct and Subject Surrogacy, *Journal of Economic Behavior & Organization*, 43(4), 499-522.

Schneider, H. S., 2012, Agency Problems and Reputation in Expert Services: Evidence from Auto Repair, *Journal of Industrial Economics* 60(3), 406-433.

Siegel, S. and Harnett, D. L., 1964, Bargaining Behavior: A Comparison Between Mature Industrial Personnel and College Students, *Operations Research*, 12(2), 334-343.

Wolinsky, A., 1993, Competition in a Market for Informed Experts' Services, *RAND Journal of Economics 24*, 380-398.

Appendix

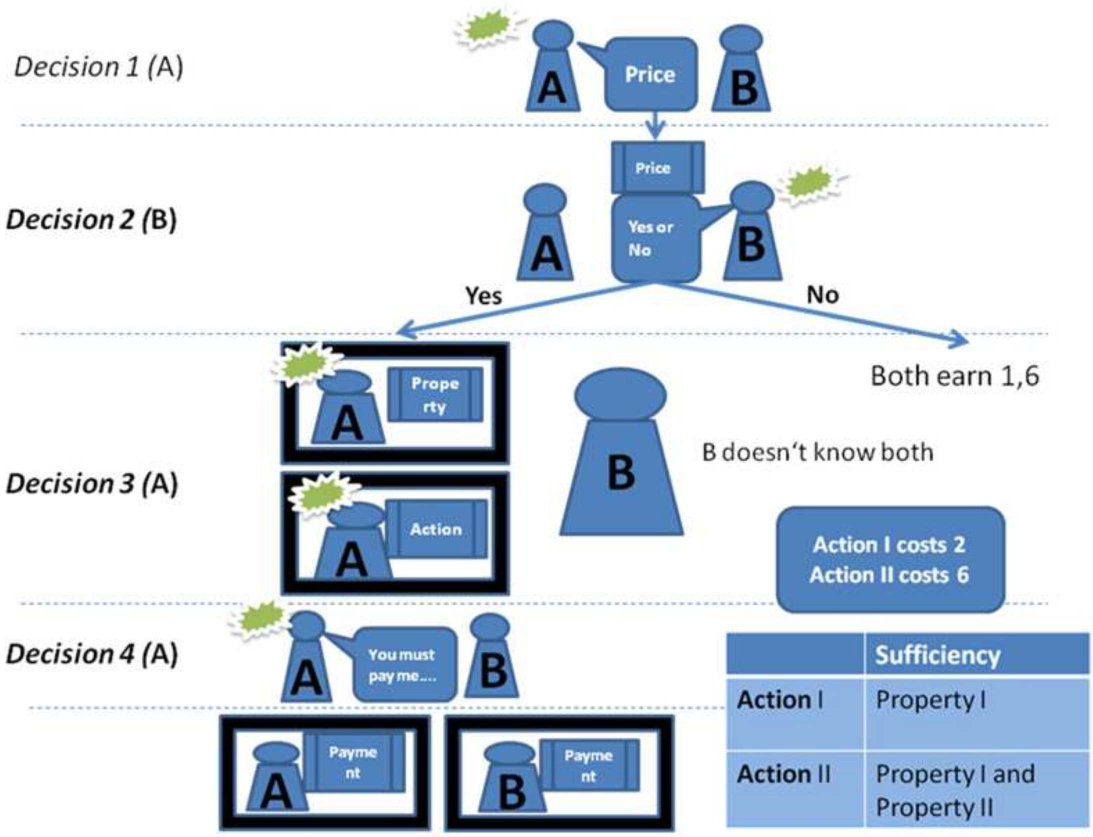

Figure 2: Example Slide Used in Car Mechanic Treatments

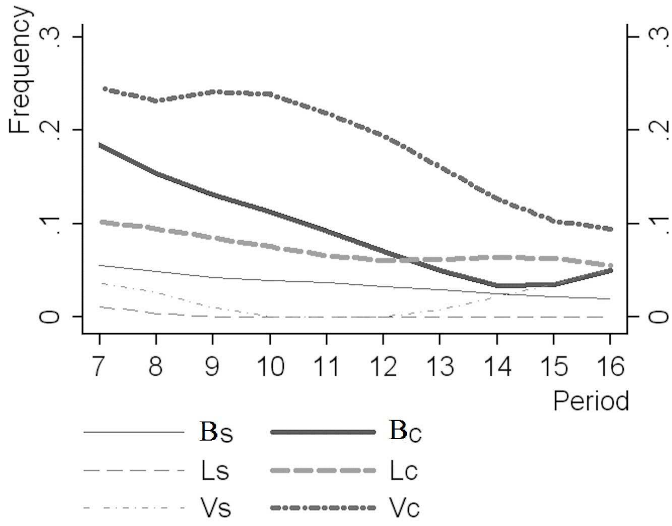

Figure 3. Overcharging over Time

Table 3. Predictions on Trade and Pricing-, Provision- and Charging-Policy

(Symmetric Perfect Bayesian Equilibrium with Own-Money Maximizing Players)

	Interaction	**Price Vector**	**Provision**	**Charging**
Baseline (**B**)	0%.	n.d.	always low	overcharging
Verifiability (**V**)	100%	(6, 10)	always low if $P^H - c_H < P^L - c_L$ efficient prov. *if* $P^H - c_H = P^L - c_L$ always high if $P^H - c_H > P^L - c_L$	honest charging
Liability (**L**)	100%	(n.d., 8)	efficient provision	overcharging

Note: n.d. stands for not determined in theoretical solution, but P^L has to satisfy $P^L \leq P^H$.