Junior Management Science 5(4) (2020) 410-428

Junior Management Science

journal homepage: www.jums.academy

On the Analysis of Moral Hazard Using Experimental Studies

Maria Huber

Ludwig-Maximilians-Universität München

Abstract

The term moral hazard generally implies individuals' tendency to exercise less effort into cost reduction if the negative consequences resulting from their actions are not borne by themselves. This paper analyzes using recent experimental studies under which circumstances moral hazard is likely to occur and how this problem could be mitigated or eliminated. A detailed overview and analysis of field and laboratory experiments from different areas are provided. At first, a description of the experimental process is presented. The paper then concentrates on findings and, additionally, on the discussion of the ethodology. Overall, the results suggest moral hazard to be an important problem in many markets. However, it is found out that experts without personal financial incentives do not respond to customers' insurance status. Besides, competition mitigates moral hazard on the supply side and evidence shows that moral hazard is less likely to occur in markets for natural disaster insurance where probabilities of damages are low. Additionally, peer pressure and pro-social preferences alleviate the problem of moral hazard in group schemes.

Keywords: First-degree moral hazard; second-degree moral hazard; experiments; analysis.

1. Introduction

Moral hazard is an issue that can occur in many different areas, but since information plays an important role, especially in insurance markets, numerous studies concerning moral hazard focus on those (Richter et al., 2014). For instance, evidence from empirical studies in the context of health insurance indicates a strong positive correlation between health insurance coverage and health expenditures while different possible explanations for this finding exist (Kerschbamer and Sutter, 2017): On the one hand, high-risk individuals are more likely to purchase insurance which refers to the problem of adverse selection. On the other hand, insured individuals demand more services or more expensive ones since their out-of-pocket costs are lower with insurance, a problem known as first-degree moral hazard. Another explanation is that physicians provide more services than necessary or more expensive ones to insured patients who are assumed to care less about the costs because costs are covered by insurance, a second-degree moral hazard problem. In order to analyze moral hazard, it is inevitable to differentiate between adverse selection, first-degree and second-degree moral hazard since the phenomena are equivalent in terms of final outcomes, but the underlying mechanisms are different (Balafoutas et al., 2017). Therefore, this paper analyzes by means of recent experimental studies under which circumstances moral hazard emerges and which features mitigate or eliminate this issue completely.

The first section of the main part contains experiments on second-degree moral hazard i.e., supply side responses to first-degree moral hazard (Balafoutas et al., 2017). Particularly, it is investigated how reimbursement by a third party affects service providers' behavior and whether eliminating financial incentives for providers or allowing for competition influences the degree of moral hazard. The presented studies concentrate on markets for credence goods since moral hazard is assumed to be specifically relevant in such markets because of the high degree of informational asymmetry between expert sellers and consumers (e.g., markets for repair services or health care). Only experts know which quality of service is needed while customers can only observe ex post whether the problem is solved, but if so, they cannot be sure of having received adequate treatment (Balafoutas et al., 2017; Dulleck and Kerschbamer, 2006). It is assumed that this information asymmetry creates strong incentives for service providers to overtreat, undertreat and overcharge (Kerschbamer et al., 2016). Especially, if providers know that consumers do not bear the costs and are, consequently, less price sensitive. Overtreatment (or overprovision) means that sellers provide higher quality or quantity of the service than

DOI: https://doi.org/10.5282/jums/v5i4pp410-428

needed to solve the customers' problem (e.g., taxi drivers taking passengers on detours) while undertreatment (or underprovision) relates to a situation where the service is insufficient (Kerschbamer et al., 2016). Overcharging is a case where experts charge for more than actually provided (e.g., computer repair experts charging the replacement of a module which has not been replaced) (Kerschbamer and Sutter, 2017). The results from the experiments show that experts without personal financial incentives did not respond to customers' insurance status (Lu, 2014). In addition, competition mitigated moral hazard on the supply side (Huck et al., 2016).

The second part of the paper concentrates on first-degree moral hazard i.e., individuals' tendency to exercise less effort if the negative consequences resulting from their actions are not borne by themselves (Balafoutas et al., 2017). It is investigated whether moral hazard exists in a market for natural distaster risk insurance. As in the case of second-degree moral hazard, first-degree moral hazard has not only been observed in insurance markets, but also in many different areas such as credit and labor markets. For instance, a person working in a team can free ride and trust on the other team members' performance when individuals are paid according to the team output (Holmstrom, 1982). Therefore, also experiments on joint liability group schemes are discussed. As a result, evidence suggests that moral hazard is less likely to occur in markets for natural disaster insurance where probabilities of damages are low (Mol and Botzen, 2018). In addition, experimenters found out that peer pressure and prosocial preferences alleviate the problem of moral hazard in group schemes (Corgnet et al., 2013; Biener et al., 2018).

The remainder of this paper is organized as follows: The next section briefly defines the term "moral hazard", explains the different types and distinguishes this problem from adverse selection. The aim of section 3 is to analyze moral hazard by using experimental studies from different areas. A detailed overview[1] of recent field and laboratory experiments is provided, due to structural reasons, first on second-degree and second on first-degree moral hazard. A description of the experimental process is for each experiment presented at first. The paper then concentrates on the results and, additionally, on the discussion of the experimental methodology. Finally, section 4 draws a conclusion and points out possible academic voids which can guide to future research topics.

2. Moral Hazard in Theory

2.1. Definition

The term "moral hazard" has its origin in the insurance literature. Arrow (1963, p. 961) defined it in the context of health insurance as the observation that "medical insurance increases the demand for medical care". Therefore, moral hazard can be viewed as an insurance-induced behavior modification of individuals (Karten et al., 2018) – meaning that an individual with more insurance coverage has weaker incentives to prevent losses and therefore insured events will occur more often compared to an individual with less or no coverage (Balafoutas et al., 2017). However, since moral hazard is not only an issue in insurance markets, the term generally implies individuals' tendency to exercise less effort into cost reduction if the negative consequences resulting from their actions are not borne by themselves (Balafoutas et al., 2017). The phenomenon of moral hazard – as of adverse selection[2] – arises from an asymmetry of information between contracting parties (Holmström, 1979; Arnott and Stiglitz, 1991). Specifically, this asymmetry occurs ex post[3] (Finkelstein and McGarry, 2006) and is associated with hidden action i.e., the probability distribution of observable outcomes is dependent on agents' actions which are unobservable for the contracting party (Arnott and Stiglitz, 1991).

2.2. Types of Moral Hazard

According to the literature on insurance theory (e.g., Nell, 1998), moral hazard can be classified into several types which are represented in Figure 1. At first, it can be divided into the legal and the illegal moral hazard. The illegal type is insurance fraud which requires a material misrepresentation (e.g., lie or concealment), the intention to deceive and the aim to realize unauthorized benefits (Viaene and Dedene, 2004).[4]

The legal type is subdivided into the external and the internal moral hazard. External moral hazard (second-degree moral hazard) references to third parties who may change their behavior based on their customers' insurance coverage whereas the internal type corresponds to the insured individuals' behavior (first-degree moral hazard) (Karten et al., 2018). The former is defined as the supply side's tendency to increase the price or the extent of a service when moral hazard on the demand side is expected since the demand side is less price sensitive due to insurance (Balafoutas et al., 2017). The second legal form includes ex ante and ex post moral hazard: Ex ante moral hazard refers to an insured individual's behavior to spend less effort in reducing the likelihood of a loss (Einav and Finkelstein, 2018; Ehrlich and Becker, 1972). For instance, an individual with health insurance coverage may have fewer incentives to avoid an unhealthy lifestyle (e.g., smoking) since insurance covers the resulting financial costs. The degree to which a subject's demand for healthcare is influenced by the out-of-pocket price he has to pay for the care is described as ex post moral hazard (Pauly, 1968; Einav and Finkelstein, 2018) i.e., if an uninsured person would not have visited a doctor because of an innocuous disease, but he decided to do so because he was insured then his behavior is attributed to ex post moral hazard. Einav and

[1] Due to space constraints, it is not possible to present a broader overview in this paper since the research question requires a detailed discussion of experiments.

[2] The problem of adverse selection is briefly addressed in section 2.3.

[3] "Ex post" relates to the conclusion of the (insurance) contract.

[4] The exact specification may vary between different systems of justice (Viaene and Dedene, 2004).

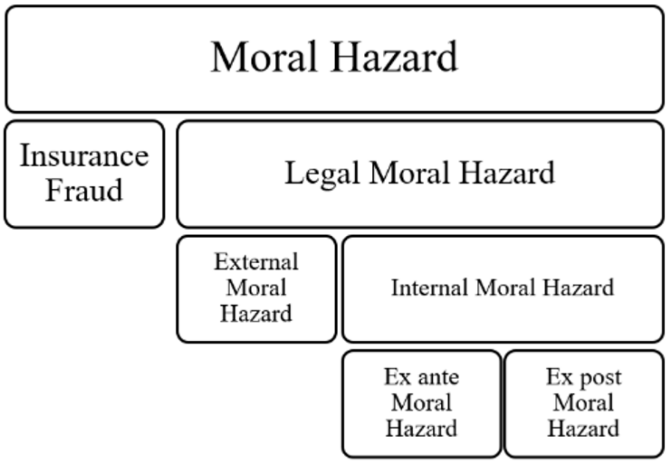

Figure 1: Types of Moral Hazard According to Nell (1998)[5]

Finkelstein (2018) argue that using "moral hazard" in this context is an abuse of the term since an individual's healthcare utilization (action) can be observed which means that there is more a problem of hidden information about the person's health status than a problem of hidden action.

2.3. Distinction from Adverse Selection

As already mentioned beforehand, the situation of asymmetrically distributed information can also lead to the problem of adverse selection which Arrow (1986) attributes to hidden information. Under adverse selection, a subject is assumed to have private information about his risk type prior to the insurance contract relative to the insurance company which creates an ex ante information asymmetry (Finkelstein and McGarry, 2006). A person with private information that he is a high risk is more likely to choose an insurance contract with a higher coverage level than a person who believes himself to be of a type of low risk (Finkelstein and McGarry, 2006). And consequently, the causality between coverage and riskiness is reversed compared to moral hazard (Finkelstein and McGarry, 2006). A positive correlation between the level of insurance coverage and the degree of riskiness can, therefore, result from both, adverse selection and moral hazard (Finkelstein and McGarry, 2006). This brings up difficulties in clearly disentangling these two problems empirically. However, this paper does neither concentrate on the analysis of adverse selection nor on approaches to clearly disentangle[6] moral hazard and adverse selection since the experiments presented in the following were designed in a way so that the problem of adverse selection was eliminated.

3. Experimental Evidence on Moral Hazard

3.1. Second-Degree Moral Hazard

Balafoutas et al. (2017) were the first to study moral hazard and its influence on market outcomes in a controlled field experiment concentrating on the effect of first-degree moral hazard on the behavior of the supply side. The authors provide evidence for second-degree moral hazard in a market for taxi rides where costs were reimbursed by a third party.

In the experiment, four research assistants, two men and two women, took undercover taxi rides in the capital city of Greece following a fixed script and secretly documented the drivers' driving and charging behavior. The rides were organized in quadruples meaning that all four assistants took a taxi from the same origin to the same destination in one or two-minute intervals and at random order. Overall, the experiment consisted of 400 rides while 200 were part of the control treatment (CONTROL) and 200 were assigned to the treatment with insurance (INS)[7]. The assistants explained to the taxi drivers in both treatments that they were not familiar with the city in order to create an information asymmetry. In CONTROL, the assistants asked the drivers shortly after the ride had begun for a receipt at the end of the ride (without mentioning the purpose of this question) while in INS, it was added that the receipt was needed since expenses were reimbursed by the passengers' employers.[8] At the end of the experiment, the actual fares paid by the assistants were compared to the correct prices. This was possible because charging fees for taxi rides are standardized in Greece: The tariff consists of a fixed fee per ride and a variable part. This variable part is either computed distance or duration-dependent contingent on what is more profitable for the driver and the

[5] Own representation based on Nell (1998).

[6] Cohen and Siegelman (2010) discuss three approaches to the disentanglement in their paper.

[7] Reimbursement from the employer and insurance have comparable financial consequences for the consumer (Kerschbamer and Sutter, 2017). Therefore, and for consistency reasons, the treatment will be declared as an insurance treatment.

[8] The authors state that, except for this additional information, both treatments were identical.

taximeter always applies the more profitable method automatically.

Balafoutas et al. (2017) measured overtreatment along two dimensions: The duration of the ride and the distance driven. Table 1 shows the values of duration and distance indices by gender and treatment. A comparison of the average duration index across treatments (1.14 in CONTROL and 1.13 in INS) and across genders (1.13 for male and 1.14 for female passengers) did not reveal any significant differences. In addition, the differences in values of the distance index were again insignificant across genders (1.06 for males and 1.07 for females), but marginally significant between CONTROL (1.06) and INS (1.08). Therefore, only a minor overtreatment effect along the distance dimension was found. The authors state that the reasons for the small differences in the overtreatment index between treatments could have been that overtreatment was associated with additional costs of service as for example fuel costs or opportunity costs of time. A passenger who does not have to bear the costs for the taxi ride would probably not mind a higher price but would complain about the duration of the ride longer than necessary.

In Table 2 one can observe overcharging frequencies and price indices across treatments and across genders. In CONTROL, 20% of taxi rides were overcharged while the overcharging frequency was 37% in INS. According to the authors, this indicates a statistically significant and causal effect of second-degree moral hazard. Additionally, the mean overcharging amount by which taxi drivers increased the fare was higher in INS (€ 1.43) compared to € 0.91 in CONTROL. Therefore, it is not surprising that the price index increased after the moral hazard manipulation as can be observed from Panel (b) in Table 2. This suggests that passengers' expenditures increased compared to the absence of second-degree moral hazard. The source of these results could be the drivers' assumption that, when employers reimburse the costs for the ride, passengers care less about higher costs and hence overcharging behavior will be undetected and not reported.

Another important finding was that in CONTROL, female passengers paid overcharged prices more frequently (26%) than male passengers (13%)[11] while the values were almost similar across genders (36% and 37%, respectively) in INS. Therefore, the difference in overcharging frequencies between both treatments was highly significant only for male passengers. Women could have been perceived as less likely to complain about being overcharged in general which explains the differences in overcharging between genders in CONTROL. And since the additional information for the driver about the employer paying for the ride did not change this perception about women, the overcharging frequency increased only by an insignificant amount of 10 percentage points.

Passengers had to pay unjustified surcharges (e.g., transport to the airport) in 77% of all overcharging cases. In the remaining ones, drivers manipulated their taximeters, applied the night tariff during daytime or rounded up the price by more than 5% of the fare. Overall, the experiment stresses the importance for employers to reduce the extent of second-degree moral hazard. As one possible solution, vertical integration with service providers (e.g., firm's own chauffeur service) may be implemented.

In the following, the experimental methodology will be discussed. List (2006) argues that field experiments represent a bridge between laboratory and naturally-occurring data. In relation to an experiment in the laboratory, the experimenter potentially has less control over the environment in a field experiment since it is not possible to take all situational factors into account (Richter et al., 2014), but in exchange for more external validity i.e., realism increases (List and Reiley, 2008). In the presented experiment, taxi drivers were the population of interest being observed in their natural environment without knowing that they were being analyzed. This is important since different types of subjects may behave differently i.e., students in the laboratory may behave differently than real taxi drivers and subjects knowing that they are being observed may also change their behavior (List and Reiley, 2008). Another advantage of the methodology was that first-degree moral hazard and adverse selection can be excluded as sources for the found results since the assistants' behavior was exogenously controlled and held constant by exact instructions and passengers were randomly assigned to one of the two treatments (Balafoutas et al., 2017). An additional benefit was that all four assistants took taxis in short intervals from the same origin to the same destination in order to make the prices comparable. Thus, factors influencing the taxi driver's choice of route (and thereby the price) as, for instance, traffic or weather conditions were eliminated (Balafoutas et al., 2017). It is important to note that the results from this experiment may not represent results in other markets (or other countries) since the market for taxi rides is highly specific and the experiment was conducted only in the city of Athens.

Due to the fact that the market for taxi rides in Greece is highly regulated and over-treatment and overcharging may have different consequences for consumers in other markets, Kerschbamer et al. (2016) confirm the importance of second-degree moral hazard in a less specific market, the computer repair market. For that purpose, the impact of customers' insurance coverage on computer repair experts' provision and charging behavior was examined.

In the natural field experiment by Kerschbamer et al. (2016), equally manipulated computers were brought to 61 randomly selected repair shops in Austria for a reparation. One of the random-access memory modules was destructed in all computers which caused an unambiguous error

[9] Balafoutas et al. (2017, p. 9); The columns CTR and MOH represent results from CONTROL and INS, respectively.

[10] Balafoutas et al. (2017, p. 10); The columns CTR and MOH represent results from CONTROL and INS, respectively.

[11] According to the authors, women were, ceteris paribus, 18.1% more likely to face overcharging in CONTROL than men. This is shown in column 2 in Appendix 1.

Table 1: Overtreatment Indices[9]

Notes. Panel (a): The duration index is the ratio of time driven in each ride to time driven in the quickest ride in that particular quadruple. Panel (b): The distance index is the ratio of distance driven in each ride to distance driven in the shortest ride in that particular quadruple. CTR refers to the control treatment and MOH refers to the moral hazard treatment.

	CIR	MOH	Overall average
Panel (a): duration index			
Male passengers	1.124	1.135	1.130
Female passengers	1.152	1.126	1.139
Overall average	1.138	1.130	1.134
Panel (b): distance index			
Male passengers	1.056	1.071	1.064
Female passengers	1.053	1.084	1.068
Overall average	1.055	1.077	1.066

Table 2: Overcharging Frequency and Price Index[10]

Notes. Panel (a): Overcharging frequency refers to the share of rides that have been classified as cases of overcharging. In parentheses, we report the mean unconditional overcharging amount (which is zero if overcharging has not taken place). Panel (b): The price index is the ratio of total price paid in each ride to the lowest total price paid in that particular quadruple. CTR refers to the control treatment and MOH refers to the moral hazard treatment.

	CTR	MOH	Overall average
Panel (a): overcharging frequency (mean overcharging amount in parentheses, in €)			
Male passengers	0.13 (0.72)	0.37 (1.46)	0.25 (1.09)
Female passengers	0.26 (1.10)	0.36 (1.40)	0.31 (1.25)
Overall average	0.20 (0.91)	0.37 (1.43)	0.28 (1.17)
Panel (b): price index			
Male passengers	1.075	1.153	1.114
Female passengers	1.109	1.177	1.143
Overall average	1.092	1.165	1.129

message on the screen.[12] Therefore, every computer expert should have been able to diagnose and solve the problem. According to the IT department, the repair should have been completed within half an hour and for costs of € 60 to € 80. The customer, an undercover experimenter, entered the shop, asked for a repair and indicated that he was a non-computer expert by mentioning that he had no idea why the computer cannot be booted. Two different treatments were randomly assigned to the shops: In CONTROL, the customer explained before leaving the shop that he would need a bill after the repair while in INS, the customer added that the bill was needed for his insurance company because repair costs were covered.[13] After the reparation, the computers were checked in order to find out what had been done to solve the booting problem and whether the positions on the bill fit to the repair actually undertaken. Finally, to investigate the motives for the differing behavior of sellers between treatments, the authors conducted a survey where they asked experts from 15 repair shops why insurance might lead to higher prices for customers.

The authors found out that the average price for the repair increased by 83% from € 70.17 in CONTROL to € 128.68 in INS indicating a highly significant effect of the insurance treatment. Figure 2 illustrates this large difference by means of the relative cumulative frequencies of repair prices. This finding is in line with what Balafoutas, Kerschbamer and Sutter found in the previous experiment.

Overtreatment yielded 29% of the price difference between the two treatments: In five cases, unnecessary repairs – additional to the replacement of the defective module – were carried out. The price for these five repairs was € 200.58 on average which was significantly higher than the average price for the other repairs in INS (€ 112.34). Interestingly, all these repairs were made in INS and since the computers were, except for the manipulation, in perfect con-

[12]According to the authors, the computers were in perfect condition aside from the manipulation.

[13]Both treatments were completely identical except for this difference (Kerschbamer et al., 2016).

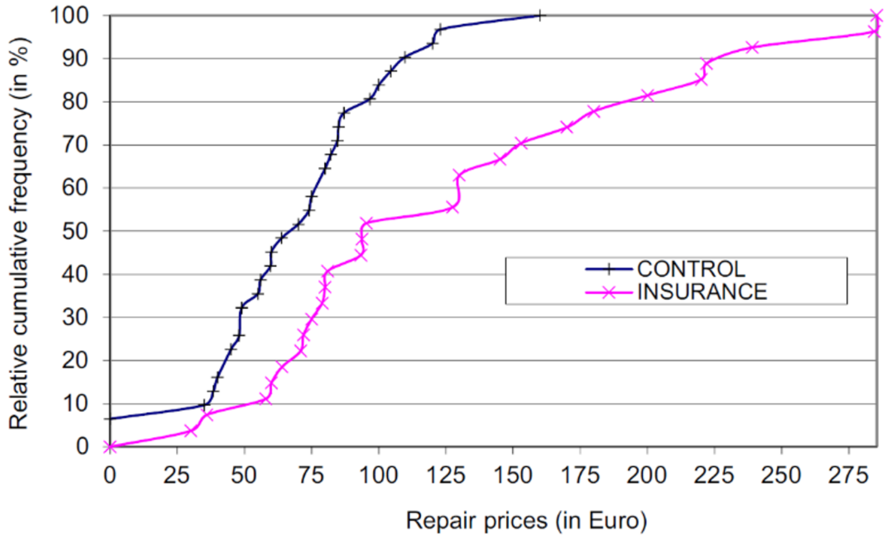

Figure 2: Relative Cumulative Frequency of Repair Prices[14]

dition, this can be interpreted as overtreatment.

Moreover, overcharging explained the remaining 71% of the price difference between CONTROL and INS: The authors found no difference due to charged repairs that had not been conducted,[15] but overcharging in the working-time dimension was found – probably since the customer was not present during diagnosis and repair. While no significant difference in hourly rates between treatments (€ 87.47 on average) occurred there was an increase in the charged working time of 85% from CONTROL (0.55h) to INS (1.02h).[16] This strong difference is also shown in Figure 3. The results from the survey on the motives for overcharging and overtreatment in the light of insurance – which are represented in Appendix 2 – showed that second-degree moral hazard was considered as the most likely explanation. Experts expected the customers to pay less attention to price minimization because of their insurance coverage.

With reference to the methodology it should be mentioned that just as in the experiment by Balafoutas, Kerschbamer and Sutter, first-degree moral hazard and adverse selection were ruled out in the experiment.[17] The computers were manipulated in a way that experts should have been able to easily find the problem and, therefore, incompetence was excluded as a reason for the differing behavior of experts. However, three shops out of 61 stated either that the computer was irreparable or that a repair would be more expensive than buying a new computer suggesting that finding the error was probably not as simple as expected. This may be an issue because when experts spend more time on identifying the problem, the repair costs increase, consequently, due to incompetence and not because of intended misbehavior (Kerschbamer et al., 2016). Therefore, it is possible that parts of the overcharging effects in the working-time dimension were solely attributed to incompetence. Another problem may be that only 29 of all repair shops indicated the working time and the hourly rate as a position on the bill. It was not possible to observe whether the charged time was used for repair or not, but it may be arguable that the number of 27 observations (two excluded because of overtreatment) was too small for drawing a justifiable conclusion on the overcharging behavior. However, a positive course of action was that observations implying overtreatment were excluded from computing the effect of overcharging in the working-time dimension since the replacement of additional parts of the computer is positively correlated with the duration of the working time. In regard to the survey, one may criticize the number of interviewed experts.

In the previously discussed experimental studies, the sellers of credence goods had financial incentives for behaving fraudulently. In contrast, Lu (2014) investigated whether experts without personal monetary incentives also react to consumers' insurance status.

The author conducted a field experiment in which undercover testers visited doctors at hospitals in Beijing (China). These testers explained to the doctors that they were sent on the authority of a family member (patient) living in another region of the country who wanted a doctor in a high-rated hospital to have a look on his case.[18] Therefore, two hypothetical patients were designed and the testers brought their reference sheets with medical test results indicating

[14] Kerschbamer et al. (2016, p. 7456)

[15] In four cases, the experts billed replacements of parts which had actually not been replaced, but two of these cases occurred in CONTROL and two in INS (Kerschbamer et al., 2016).

[16] Since only 29 shops indicated the rate per hour and working time on the bill (and two observations were excluded due to overtreatment), these numbers were computed from 27 observations.

[17] See the discussion of the methodology of the previous experiment for an explanation.

[18] This procedure is very common in China (Lu, 2014).

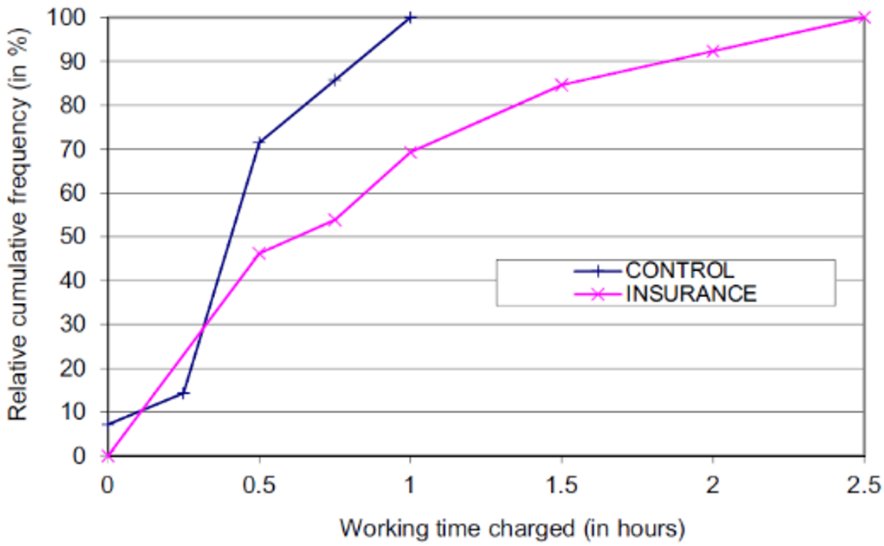

Figure 3: Relative Cumulative Frequency of Working Time[19]

different health problems which either required medication or not. After the tester had described the patient's health problems according to a fixed script, the doctor had to decide whether to prescribe no drugs, generic drugs or more expensive brand-name drugs and the package size. The experiment was divided into four different treatments which were randomly assigned to the doctors: Insurance-incentive, no-insurance-incentive, insurance-no-incentive and no-insurance-no-incentive. In the treatments with insurance, doctors received the information that the patient was insured and, to the opposite, that he had no insurance coverage in the no-insurance treatments. Additionally, in the incentive treatments, doctors were informed that the tester would buy the prescribed drugs for the patient from the doctor's hospital. This case created personal financial incentives for physicians since their payments often depend on the revenue generated in their hospital.[20] The testers indicated that the drugs would be purchased elsewhere in the no-incentive treatments.

Evidence presented in Table 3 shows that when doctors had financial incentives to prescribe more drugs or more expensive ones, patients paid 522 Yuan on average in the insurance condition and 365 Yuan when they were not insured. Therefore, insured patients had to pay 43% more for drugs – which was highly statistically significant – since physicians prescribed more brand-name drugs (83% vs. 68%), a higher number (2.47 vs. 2.20) and more units of drugs (2.53 vs. 2.09) to insured.

These effects are displayed in the first column in Appendix 4. A possible reason for these results could have been that doctors wanted to increase drug expenditures since their income was calculated in proportion. An important finding was that in the no-incentive treatments average outcomes were not statistically different between insurance statuses as can be seen from the second column in Appendix 4. Hence, physicians did not respond to the patients' insurance status when they did not expect any profits from prescriptions. By comparing the insurance-incentive to the insurance-no-incentive treatment, the author found out that doctors with incentives prescribed significantly more unnecessary drugs[21] to insured patients (second line in Table 3). The number and units of drugs were also significantly higher, but the share of branded drugs was almost equal in both treatments (83% and 81%). Overall, given insurance, financial incentives for doctors increased the average drug expenditures for patients significantly.

In this experimental study, adverse selection and first-degree moral hazard did not play a role either since testers, who were randomly assigned to the treatments, received exact instructions for their behavior. In addition, the testers indicated that they were not the patient who needed the doctor's advice and, therefore, the testers' characteristics should have had little impact on the doctors' behavior (Lu, 2014). However, Lu (2014) does not completely rule out the possibility that doctors' inferred information from the conversation with the tester may have influenced the results. For instance, although the author implemented two elements to make the doctors aware of the patients being neither poor nor rich, the doctors could have assumed that patients who did not want the drugs to be purchased in the hospital were more price sensitive since – according to Lu (2014) – prices

[19]Kerschbamer et al. (2016, p. 7456)

[20]In addition, hospitals in China often receive kickbacks from drug companies which also results in incentives for doctors to prescribe (Yip and Hsiao, 2008).

[21]One hypothetical patient had increased triglycerides, but, according to medical guidelines, the patient should have not received medication for this level of triglycerides (Lu, 2014). Therefore, it was possible to test for overtreatment.

Table 3: Average Treatment Outcomes[22]

Notes: "D&H" represents "for diabetes and hypertension only." Standard errors are in parentheses.

Dependent variables	Insurance incentive	No insurance incentive	Insurance no incentive	No insurance no incentive
For both patients				
Raw drug expenditure (Yuan)	522.11	365.14	-	
s.e.	(35.80)	(23.63)	-	
Prescription for triglycerides (0/1)	0.64	0.40	0.28	0.40
s.e.	(0.10)	(0.10)	(0.09)	(0.10)
Monthly drug expenditure D&H (Yuan)	424.78	298.71	324.50	307.03
s.e.	(23.54)	(15.84)	(18.95)	(15.44)
Number of drugs D&H	2.47	2.20	2.18	2.18
s.e.	(0.10)	(0.08)	(0.07)	(0.06)
Unit of drugs D&H	2.53	2.09	2.16	2.12
s.e.	(0.11)	(0.08)	(0.09)	(0.07)
Share of branded drugs D&H (0-1)	0.83	0.68	0.81	0.80
s.e.	(0.04)	(0.05)	(0.03)	(0.04)
Obs. for triglycerides	25	25	-	-
Obs. for other variables	49	49	49	49

at outside pharmacies can be below the ones in hospitals. Besides, it would have been optimal to visit each doctor for all four treatments to control for heterogeneity between the doctors, but Lu (2014) argument that presenting the same test results multiple times to each doctor would have caused suspiciousness among physicians seems plausible.

The first laboratory experiment discussed in this paper was conducted by Huck et al. (2016) who investigated the effects of medical insurance and competition on patients' and physicians' behavior with a focus on overtreatment. They did not only find second-degree moral hazard, but also evidence for first-degree moral hazard. The experiment consisted of four treatment variations – CONTROL, INS, treatment with competition (COMP) and treatment with insurance and competition (INS-COMP) – which are explained in the following.

In CONTROL, 336 students were randomly assigned to a fixed role as a physician or as a patient. The patients were confronted with a problem which required treatment. In each round, the patients were randomly matched to a physician and the severity of their problem (mild or severe) was determined. Then, patients had to choose – without knowing the severity of their problem – whether to consult their assigned physician or not. If a patient consulted a physician, he was able to observe the severity of the problem and chose the treatment (patients had to pay for treatments). In the case of a severe problem, the physician only had the option to provide a severe (and more costly) treatment to the patient as presented in Figure 4 whereas in case of a mild problem, the physician also had the opportunity and monetary incentives to overtreat. This means offering a severe treatment although a mild treatment would have been sufficient for a cure and less expensive for the patient. The payoffs resulting from each condition can be observed from parentheses in Figure 4: The upper numbers are patients' payoffs while lower numbers are those of physicians. At the end of each round, patients who consulted their physician received information about their treatment, but still not about the severity of their problem. This type of information was only given to those who decided not to consult. All subjects also learned about their payoffs after each round.

INS was almost equal to the above-described process, but all patients – even the ones not consulting a physician – had to pay a fair premium to cover extra costs of overtreatment. An additional difference was that prices for severe and mild treatments were – in contrast to CONTROL – identical. Hence, a single patient did not have to bear higher costs resulting from unnecessary overtreatment alone. The premium was calculated dependent on the number of severe treatments provided i.e., the more severe treatments provided the higher the premium. Patients were aware of this calculation method and received information about the paid premium at the end of the rounds.

In COMP, patients were allowed to freely choose among all physicians which was defined as competition. In addition, patients and physicians observed the number of patients who had consulted the physician in previous rounds (i.e., market shares) from a history table. Finally, in INS-COMP, INS and COMP were combined to one treatment.

Huck et al. (2016) found out that insurance induced moral hazard on both sides of the market. Table 4 which summarizes average results from all rounds and for all treatments shows that 36% more patients consulted a physician in INS than in CONTROL because additional costs of treat-

[22] Lu (2014, p. 161); Due to space limitation, the complete table cannot be presented in this part of the paper, please see Appendix 3.

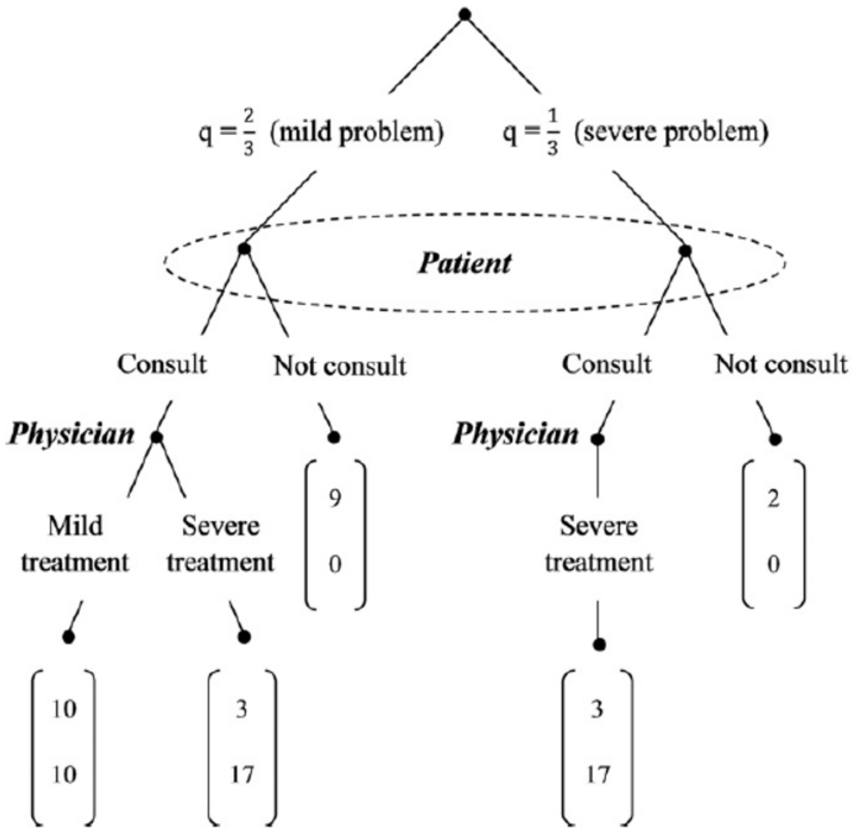

Figure 4: Game Tree of CONTROL with Actual Payoffs[23]

Table 4: Results from All Treatments[24]

Notes: The table shows averages over all 30 periods and 7 markets in the main treatments. The rates in the first four lines are indicated in percent: (1) is the share of consulting patients, (2) is the share of consulted physicians who give severe treatment when the problem is mild, where the average rate (2) is weighted by the number of consultations per session and period. (3) is the sum of actual earnings over the sum of potential earnings. (4) is the share of all interactions when appropriate treatment is provided. Average earnings in (5) and (6) are indicated in points.

	BASE	COMP	INS	INS-COMP
(1) consulting rate	40.7	54.7	55.3	83.1
(2) overtreatment rate	26.3	7.2	70.9	34.2
(3) efficiency rate	61.2	70.5	71.5	89.5
(4) correct treatment rate (CTR)	29.6	49.7	16.2	54.9
(5) average earnings physicians	9.1	11.5	14.4	19.1
(6) average earnings patients	6.8	7.2	5.7	6.4

ment were paid by the collective.[25] The overtreatment rate of 70.9% was about three times the level in CONTROL (26.3%). Physicians had additional incentives to overtreat because they assumed that patients were less concerned about the costs. Overall, only 16.2% of patients received a correct treatment in INS – compared to 29.6% in CONTROL.[26] However, the effect of insurance was stronger in the context of competition: Insurance increased the consulting rate by 52% from 54.7% to 83.1% and the overtreatment rate (34.2%) was about five times as high as in COMP (7.2%).

The lowest overtreatment rate (7.2%) was measured in

[23]Huck et al. (2016, p. 85)

[24]Huck et al. (2016, p. 87); The column BASE represents results from CONTROL.

[25]This effect was not significant according to a Wilcoxon-Mann-Whitney test (Huck et al., 2016). More information on the test results is provided in Appendix 5.

[26]To measure efficiency, the correct treatment rate instead of the efficiency rate is used since the latter does not take overtreatment as an inefficiency into account i.e., efficiency is high even when all patients consulted, but all were overtreated.

COMP compared to CONTROL (26.3%), INS (70.9%) and INS-COMP (34.2%) since competition provided incentives for physicians to avoid a severe treatment when a mild treatment would have been sufficient. Overtreating physicians were less likely to be consulted and therefore under pressure not to overtreat. Probably the most important finding was that competition[27] on the seller side outweighed some of the moral hazard effects: On the one hand, competition increased the consulting rate from 55.3% in INS to 83.1% in INS-COMP. But on the other hand, competition reduced physicians' overtreatment behavior by 48% from 70.9% to 34.2% yielding almost the level in CONTROL. As a result, the correct treatment rate raised from 16.2% in INS to 54.9% in INS-COMP.

In the following section, the methodology will be critically discussed. In general, an important advantage of laboratory experiments is the ability to control most aspects of the environment, but such experiments may have limited relevance for individuals' actual behavior in real-world situations (lack of external validity) since subjects typically know that they are part of an experiment and the environment might not be fully representative (e.g., students as subjects) (Richter et al., 2014; List and Reiley, 2008). Adverse selection was ruled out because patients' problems were drawn randomly for each round and there was no option to not insure or to choose different coverage levels. According to Huck et al. (2016), the findings should only be interpreted in a healthcare context with fee-for-service remuneration systems i.e., where physicians take advantage from offering high-level treatments. Hence, one drawback may be that the authors did not frame the experiment in a medical context (e.g., physicians were called "advisers"). The authors named several reasons for doing so, but this feature may have influenced patients' consulting decision since consumers are probably more sensitive about their decision when it comes to their health rather than in other contexts (e.g., problem with their car). Moreover, another disadvantage of the experiment is that patients did not suffer from physical consequences (e.g., pain) after not consulting a physician. Feeling such negative consequences may have had a stronger impact on the patients consulting decision for the following rounds than just learning the severity of the problem. Additionally, the difference in patients' payoffs between not consulting and consulting and receiving the right treatment was very small (see Figure 4). It should be mentioned that especially in the case of a severe problem this seems unrealistic although the severe treatment was very expensive. Contrariwise, the value for a person of being cured is difficult to measure and may differ from person to person. The experiment focused on over-treatment and, therefore, undertreatment and no treatment were excluded, but both cases may occur in real situations. Furthermore, it was assumed that physicians diagnose the problem correctly which is obviously an unrealistic assumption for the real world. Reputational incentives for physicians were weak since patients only knew whether their problem had been solved but had no idea about the necessity of the treatment. However, such incentives may be important to mitigate overtreatment since patients could have been more confident not to be overtreated, especially, in the context of competition. A doctor's reputation (e.g., internet portals like sanego) may have an influence on the number of consulting patients.

3.2. First-Degree Moral Hazard

Results from the prediscussed experiment also demonstrate evidence of first-degree moral hazard. Thus, the remaining part of this paper analyzes this phenomenon. Previous studies found, for instance, that moral hazard is less likely to occur under deterministic losses (Berger and Hershey, 1994) and with low probabilities of obtaining income loss compensation (Di Mauro, 2002).

Mol and Botzen (2018) were the first to experimentally study the existence of moral hazard in a market for natural disaster risk insurance. To be more specific, the causal effects of different financial incentives, probability levels, behavioral characteristics and deductibles on homeowners' damage reducing investments were examined.

In a laboratory experiment, participants played an investment game on computers for which they were randomly assigned to five different treatments: CONTROL, INS, treatment with premium discount (DISCOUNT), treatment with loan (LOAN) and treatment with loan and discount (LOAN-DISCOUNT). In CONTROL, subjects had no insurance coverage whereas in INS, participants were covered by insurance including a deductible. All treatments, except for CONTROL, implied insurance coverage and a deductible. In DISCOUNT, subjects were offered a premium discount proportional to their investment in damage reduction. In the fourth treatment – LOAN – participants were able to distribute their investment costs over multiple rounds at an interest rate of 1%. Subsequently, LOAN-DISCOUNT combined the previous two treatments.

At the beginning of the experiment, participants earned their starting capital through an effort task in order to purchase a virtual house which was prone to flood risk. The rest of the starting capital was subjects' savings which could have been used to pay for investments, insurance premiums and damages. Altogether, participants played 6 scenarios[28] consisting of 12 rounds with differing flood probabilities and deductible levels for each scenario (see Table 5). The scenarios started with the presentation of flood probability, estimated maximum flood damage and deductible level on subjects' screens. On the next page, which is displayed in Appendix 6, subjects were asked how much they wanted to invest in reducing the damage of a flood in the following rounds. For this purpose, five investment levels were proposed each specifying the costs for the investment, the amount by which the

[27]The authors state that the strong effect of competition was due to free choice of physician rather than to observability of market shares.

[28]After each scenario, the savings were automatically restored to the starting value (Mol and Botzen (2018).

Table 5: Overview of Scenarios[29]

Insurance treatments			
	Deductible		
	Extra Low (5%)	Low (5%)	High (20%)
Low probability (3%)	LxL	LL	LH
High probability (15%)	HxH	HL	HH

No Insurance treatments						
Probability	1%	3%	5%	10%	15%	20%

damage will be reduced and the resulting deductible if the house was flooded. After the subjects' decision, the premium and the investment costs were subtracted from the savings. Then, subjects observed a map indicating 100 houses under which their house was marked. The software randomly selected the flooded house(s) according to the flood probability and indicated the flooded house(s) in blue on the map (see Appendix 7). If a subject's house was flooded, the deductible – or the damage in CONTROL – was paid from his savings. In the following round, subjects could have either invested more or remain with their investment while a reduction was not allowed. Additionally, in LOAN an extra page to pay the loan costs was displayed to subjects. In the end, questions and decision tasks were presented to participants in order to obtain their behavioral characteristics.

The authors found out that average investments in damage reducing measures increased with the expected value of damage (i.e., higher probability of flood and/or higher deductible) for CONTROL and INS which can be seen in Figure 5. In the first round of INS, average investments were greater than zero for high and low-probability scenarios. Moreover, subjects invested significantly less in INS than in CONTROL in scenarios with high probabilities (15%) while such an effect was not found for low probabilities (3%). These results suggest the existence of moral hazard in scenarios with high probabilities, but not under low probabilities. Therefore, moral hazard may be less of a problem in natural disaster insurance markets with low probabilities of loss and high expected damages.

The results from Table 6 prove that premium discounts increased investments in damage mitigation significantly compared to INS. But in LOAN, participants were not more likely to invest more than in INS. Consequently, the combination of loan and discount did not generate the highest investments as hypothesized by the authors.[30]

Furthermore, participants' behavioral characteristics such as risk aversion, perceived effectiveness of protective methods and concern about flooding had a positive impact on the investment decision. However, females invested significantly more than males and individuals with high incomes in real life invested less compared to low-income participants. It was also found that subjects who had already experienced a flood invested extra in damage reduction afterwards, but this effect was not found when a neighbor's house had been flooded.

One drawback – as in many other laboratory experiments – is that study participants were largely students who may have not been representative subjects to study moral hazard in a flood insurance context because of less knowledge and experience with damage reducing investments compared to real homeowners. This could explain the result from LOAN since students may have generally disliked lending money or were put off by the interest rate (Mol and Botzen, 2018). In addition, investment costs were distributed across 12 rounds which lasted in the experiment at most several minutes while costs are spread over multiple years in the real world (Mol and Botzen, 2018). Smith (1982) stresses that salient payoffs – rewards for individuals' participation in the experiment that are related to participants' realized outcomes – are important in laboratory experiments.[31] Such payments need to be incentive compatible i.e., payments create incentives for subjects to behave according to their real preferences (Jaspersen, 2016). Therefore, an advantage of the experiment was the implementation of an incentive-compatible payment scheme: At the end of the experiment, all subjects were paid their final savings from one randomly selected scenario additionally to a participation fee (Mol and Botzen, 2018). Another point is that, for ethical and practical reasons, it is not possible to let subjects lose money for real in an experiment (Etchart-Vincent and l'Haridon, 2011; Jaspersen, 2016). In order to solve this problem, subjects had to perform an effort task in which they earned their initial endowment from which they lost without affecting their own money. It is important in experiments to make subjects believe that the earned (and lost) money is theirs in order to make them aware of losing instead of gaining money in the game. Otherwise, subjects may keep their endowment in mind when making decisions and consider their outcomes as gains causing biases in their

[29] Mol and Botzen (2018, p. 8)

[30] Indeed, premium discounts alone led to the highest investments in the game (Mol and Botzen, 2018).

[31] Camerer and Hogarth (1999) found out that salient rewards change participants' behavior in experiments.

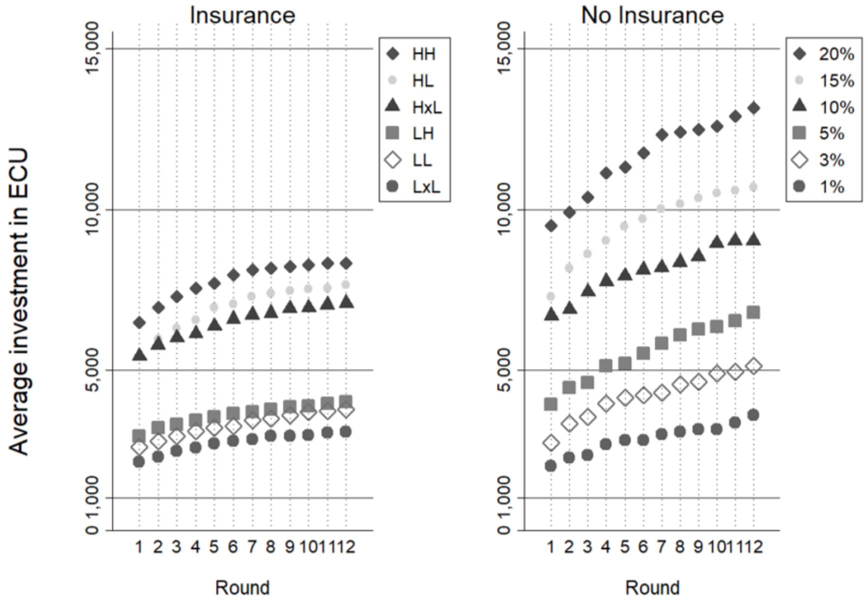

Figure 5: Average Investment in INS and CONTROL[32]

Table 6: Average Investment in the First Round for All Insurance Treatments[33]

	Baseline Insurance	Loan	Discount	Loan+Discount	Kruskall-Wallis test
scenario HH	5,421.49	3,711.86	9,233.33	8,614.04	$\chi^2 = 37.670^{***}$
	(5,431.01)	(3,658.01)	(5,732.35)	(5,512.18)	
scenario HL	4,049.59	2,847.46	8,416.67	7,807.02	$\chi^2 = 43.713^{***}$
	(4,843.98)	(3,916.43)	(5,681.64)	(5,717.89)	
scenario HxL	3,471.07	3,542.37	8,966.67	7,771.93	$\chi^2 = 46.829^{***}$
	(5,010.11)	(5,032.04)	(5,971.59)	(5,840.19)	
scenario LH	2,727.27	1,661.02	3,850.00	3,719.30	$\chi^2 = 10.086^{**}$
	(4,222.95)	(3,412.00)	(4,398.86)	(4,806.08)	
scenario LL	2,404.96	1,525.42	3,283.33	3,421.05	$\chi^2 = 10.842^{**}$
	(4,253.58)	(3,650.02)	(4,584.76)	(5,119.81)	
scenario LxL	1,793.39	1,406.78	3,550.00	2,087.72	$\chi^2 = 19.308^{***}$
	(3,976.84)	(3,312.04)	(4,560.05)	(3,434.49)	
Observations	121	59	60	58	

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$, st.dev in parentheses.

behavior (Harbaugh et al., 2010; Jaspersen, 2016).[34]

Moral hazard in teams[35] may arise when team members bear the costs of their effort for supplying inputs individually, but only the joint output is observable directly (Holmstrom, 1982). This can cause a free riding problem: Agents can cheat and rely on the performance of the remaining team members when they are paid according to the team output (Holmstrom, 1982; Corgnet et al., 2013). Corgnet et al. (2013) investigated whether peer pressure[36] through monitoring is a solution to the problem of moral hazard in teams. For this purpose, four treatments were designed: CONTROL, treatment with team incentives (T), treatment with team incentives and visible peer monitoring (TVP) and treatment with team incentives and invisible peer monitoring (TIP).

In a laboratory experiment, participants had to do a long, repetitive and effortful work task which consisted of summing up tables. When a subject had completed a table, he received information about his accumulated individual production: The production increased by 40 Cents when the table

[32] Mol and Botzen (2018, p. 17)

[33] Mol and Botzen (2018, p. 19); The column Baseline Insurance represents results from INS.

[34] However, Etchart-Vincent and l'Haridon (2011) found the contrary.

[35] Holmstrom (1982) defines a team as a group of individuals whose productive inputs are related.

[36] Mas and Moretti, 2009 define social (or peer) pressure as the experience of disutility when being observed working less hard than others.

was completely correct and decreased by 20 Cents for each incorrect given answer in the table. Moreover, at the end of each period, participants were informed about the total profit generated by their team (10 group members). Anytime during the experiment, subjects could have surfed the Internet which was an alternative activity to the work task not generating any profits. Since both activities were undertaken on different screens completing tables while browsing simultaneously was not possible. Additional to the prementioned activities, subjects could have clicked on a yellow box on their screen which was the clicking task. The box appeared every 25 seconds on the screen and by clicking on it, 5 Cents were earned by the subject.[37] As a consequence, subjects could have earned money constantly without actually working on the working task and while being on the Internet. In CONTROL, subjects had individual incentives and received payments for the work task according to their individual production whereas in T, rewards were based on the total production of all group members (subjects obtained 10% of the total production). The third treatment variation – TVP – was similar to T except for the introduction of an option for peer monitoring in order to create peer pressure. Subjects were allowed to click on a watch option on their screen to observe other participants' activities in real time. During monitoring others, the working task and the leisure activity could not have been continued while proceeding with the clicking task was possible. After selecting the watch option, monitors were informed about each subject's activities (work task, browsing or watching), production and the individual input to the work task expressed as a percentage. Additionally, monitored subjects received a notification on their screen that they were currently being watched. In TIP, participants did not receive such a notification.

The results indicated that individual production[38] per period increased (except for period 3) under individual and team incentives showing evidence of a learning effect. Subjects evolved strategies to compute the tables more quickly. Figure 6 illustrates the interesting finding that average production per subject was significantly lower in T (2.83 tables) than in CONTROL (4.21 tables) yielding a difference of 49% between the two incentive schemes due to shirking behaviors.

The following results are important since a highly significant negative correlation between Internet usage and individual production for treatments with individual and team incentives was detected: A comparison of Internet usage revealed that the average time spent with browsing was significantly higher in T (28.5%) than in CONTROL (11.9%) which can also be seen in Figure 7.[39] Under team incentives, the average proportion of time spent on the Internet of 28.5% decreased to 13.1% with the introduction of peer monitoring in TVP resulting in values almost similar to CONTROL (see Figure 7). This showed a clear impact of peer monitoring on subjects' choice of activity. Especially, visible monitoring was effective since Internet usage was significantly lower in TVP than in TIP.

Average production was 47% higher under peer pressure (in TVP) than in T which was interpreted as evidence of a strong peer pressure effect while no significant differences between TVP and CONTROL were found as shown in Figure 8. Therefore, visible peer monitoring combined with team incentives allowed production levels as high as under individual incentives supporting the authors' expectation that peer pressure eliminates the problem of moral hazard in teams. Social pressure was essential for the effectiveness of monitoring since production levels were significantly lower in TIP than in TVP and almost as high as under team incentives.

An advantage of the methodology was that subjects could have switched to the leisure activity since access to the Internet at the workplace is very common in organizations and according to a recent survey of Salary.com (see Appendix 8), 64% of employees visit websites which are not related to their working activity every day. The study also revealed that one of the most time-consuming activities employees waste their time with on the job is surfing the Internet. Corgnet et al. (2013) conducted the invisible monitoring treatment (TIP) with the objective of eliminating the role of social pressure in contrast to TVP. Yet, subjects knew about the possibility of monitoring others and may have felt watched even without a notification on their screen. Therefore, social pressure may not have been completely eliminated (Corgnet et al., 2013). This hypothesis was supported by the finding of a slight difference in Internet usage between TIP (19.8%) and T (28.5%). In the experiment, intrinsic motivation[40] was reduced through the introduction of a long and laborious work task because of the aim to investigate behavior under different incentive schemes. Corgnet et al. (2013) stated that intrinsic motivation would have been a confounding factor, but individual production may not always be driven only by extrinsic motivators such as the payment. For instance, the work itself or recognition should also be taken into consideration when conducting experiments on teamwork. Only large teams consisting of ten individuals were studied, but much work in organizations is performed by small teams. By keeping teams small it may be easier to increase the transparency of subjects' individual contribution to the output possibly even without monitoring.

In the absence of peer pressure, Biener et al. (2018) stud-

[37]This feature represented the payment an employee receives only for being present at his workplace (Corgnet et al., 2013).

[38]Production is defined as the monetary amount generated from working on the work task divided by 40 Cents. Thus, production is the number of correctly computed tables minus the number of false answers (Corgnet et al., 2013).

[39]According to the authors, 40.9% and 11.7% of subjects never surfed the Internet under individual and team incentives, respectively.

[40]Intrinsic motivation is defined as performing an activity because of the activity itself (perceived enjoyment) and not because of achieving valued outcomes (perceived usefulness) (Teo et al., 1999).

[41]Corgnet et al. (2013, p. 16)

[42]Corgnet et al. (2013, p. 26)

[43]Corgnet et al. (2013, p. 23)

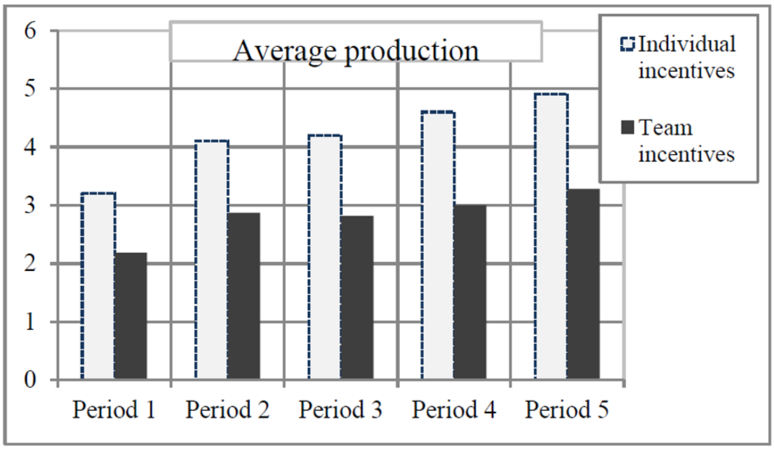

Figure 6: Average Production in CONTROL and T[41]

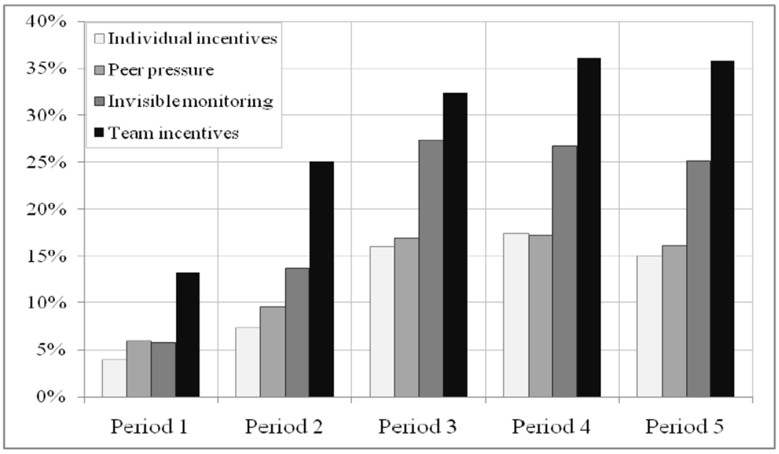

Figure 7: Average Internet Usage for All Treatments[42]

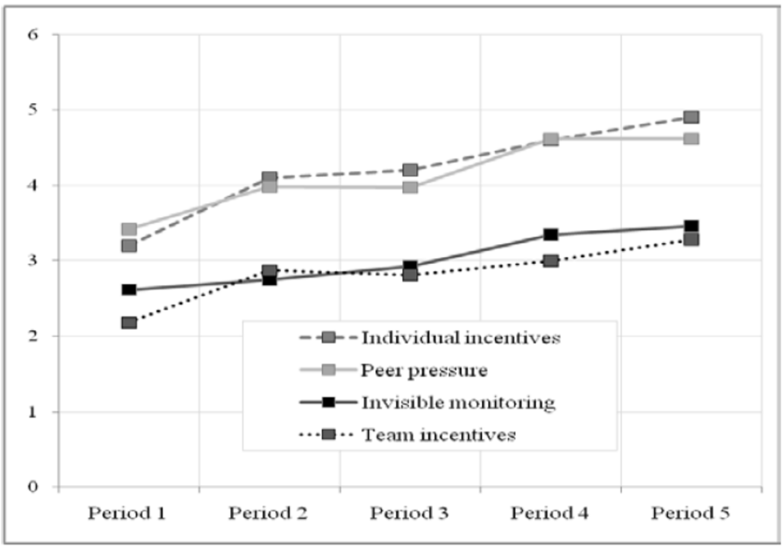

Figure 8: Average Production for All Treatments[43]

ied whether pro-social preferences[44] between agents mitigate moral hazard in joint liability group contracts. At the beginning of the experiment, all participants played the same effort game for three rounds: Subjects obtained an initial endowment[45] W and were confronted with a lottery in which they had to draw one of ten balls from a bag containing four orange and six white balls. The orange balls indicated a loss of L while white balls represented no loss. Subjects were offered the opportunity to self-protect by replacing the bag with another bag with two orange and eight white balls in exchange for effort costs of e. The probabilities and payoffs for the basic and the self-protection game are presented in Table 7.

Subsequently, subjects were randomly assigned to one of five treatments for the remaining three rounds: CONTROL, treatment with individual insurance and low coverage (Ilow), treatment with individual insurance and high coverage (Ihigh), treatment with group insurance and private information ($G_{private}$) and treatment with group insurance and public information (G_{public}). In CONTROL, subjects played the above-described game for all rounds. The insurance treatments – in which subjects paid a premium in each round in order to only bear the deductible in case of a loss – were divided into individual insurance and group insurance. The individual insurance treatments Ilow and Ihigh included different coverage levels i.e., deductibles – a high deductible leading to low and a low deductible resulting in high coverage, respectively. The group insurance treatments $G_{private}$ and G_{public} were similar to Ilow concerning deductible and premium, but insurance covered groups of two individuals and losses which were not covered by insurance (i.e., losses below the deductible) were shared among them. In $G_{private}$, self-protection was private information which means that the peer was not able to observe the other's risk-taking behavior whereas in G_{public} this information was provided at the end of the round.

The experiment was conducted two times. First, with rural villagers from the Philippines who brought a friend or a relative of choice to the experiment. Therefore, it was possible for the experimenters to manipulate the network strength within the groups. Second, in a computer laboratory setting, the behavior of students from Germany was studied. In this variation, participants could not have identified their group members.

Biener et al. (2018) found strong evidence for moral hazard meaning that subjects' effort to self-protect decreased with increasing insurance coverage, i.e., when making the payoff less state-dependent as illustrated in Figure 9. From Figure 9 one can also observe that this effect was significantly stronger in the German sample. However, moral hazard was mitigated with joint liability in the group scheme. In $G_{private}$, self-protection increased compared to Ilow by 6.8 percentage points in the Philippine sample and relative to Ihigh by 15.6 and 27.3 percentage points in the Philippine and the German sample, respectively. According to the authors, this effect was driven by individuals with positive beliefs about their peer's self-protection behavior. Subjects with positive beliefs (bar (a)) were more likely to self-protect themselves compared to individuals with negative beliefs (bar (b)) as can be seen in Figure 10. This indicates that positive beliefs about the group member increased pro-social motives and trust in the peer influenced the subject's own behavior. Also, results suggested that strategic motives (such as fear of punishment) can further improve effort provision in non-anonymous groups: Slightly higher self-protection was detected in G_{public} compared to $G_{private}$, but only for the Philippine sample. For both samples, mean self-protection for subjects with positive beliefs was almost similar in both group insurance treatments while only in the Philippine experiment, a difference was found for subjects with negative beliefs. To be more specific, the mean proportion of self-protecting individuals was 18.3 percentage points higher in G_{public}. These results are presented in Appendix 9.

Network strength, i.e., group composition, was expected to influence the degree of pro-social and strategic motives to provide self-protection. Indeed, in the Philippine experiment, whether a group was formed out of friends and relatives (strong group) or of random persons (weak group) did not cause a significant difference in outcomes of $G_{private}$ and G_{public} (see Figure 11). It was only found that the outcome difference between G_{public} and $G_{private}$ was larger in strong groups. A possible reason may be that, if self-protection behavior is observable for the peer, image concerns play a larger role in non-anonymous groups. Anyway, it is important to mention that even random individuals from the same village may have known each other.

With regard to the methodology, it can be said that one major advantage was the execution of the experiment in two independent and different environments. Once, with low-income rural villagers from the Philippines and again with students from Germany. In consideration of the fact that the main results were found for both experiments and, therefore, hold independently of culture and experience with financial services, one can attribute them further credibility, according to Biener et al. (2018). In contrast to the experiment by Mol, Botzen and Blasch, subjects did not earn their initial endowment through an effort task. Instead, they received a so-called windfall payment at the beginning meaning that participants were endowed with money without consideration (Jaspersen, 2016). According to Jaspersen (2016), the experimental design was constructed incentive compatible through using a "random problem selection mechanism". This means that subjects had to take multiple self-protection decisions during the experiment, but earnings for subjects'

[44]Meaning that individuals are also concerned with other individuals' payoffs, not only with their own ones (Levitt and List, 2007).

[45]After each round, the initial endowment was restored to the starting value (Biener et al., 2018).

[46]Own representation based on Biener et al. (2018)

[47]Biener et al. (2018, p. 241); The bar C represents results from CONTROL.

[48]Biener et al. (2018, p. 241); The bars represent the proportion of individuals choosing self-protection who had positive (a) and negative (b) beliefs about their peer's self-protection effort.

Table 7: Payoffs and Probabilities[46]

	Payoff (Probability)	
	Basic Game	Self-Protection Game
No Loss	W (60%)	W-e (80%)
Loss	W-L (40%)	W-e-L (20%)

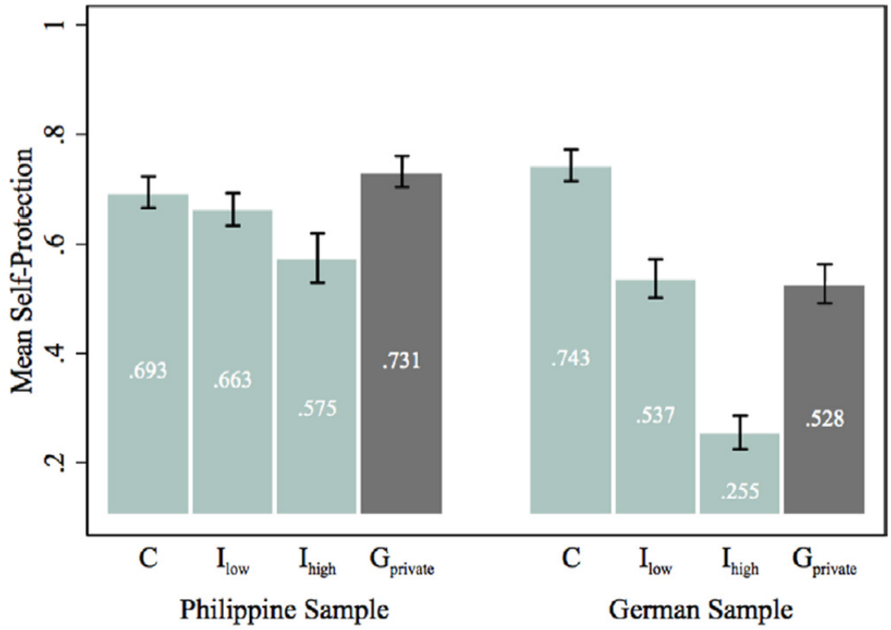

Figure 9: Mean Proportion of Self-Protecting Participants[47]

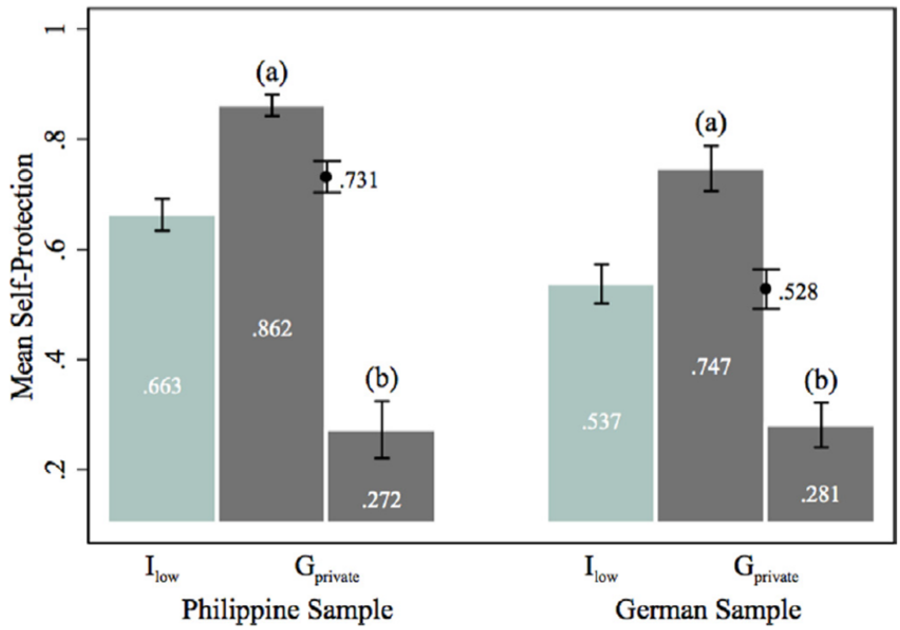

Figure 10: Mean Proportion of Self-Protecting Participants by Beliefs[48]

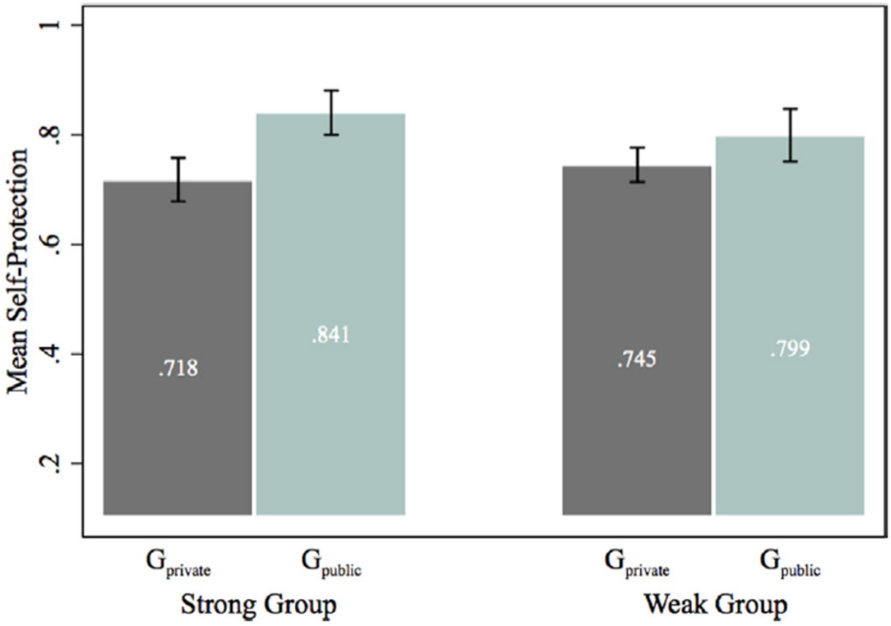

Figure 11: Mean Proportion of Self-Protecting Participants by Group Composition[49]

participation were determined by randomly selecting one round out of six (in addition to a show-up fee). Biener et al. (2018, p. 237) write in their paper that they “explained the [. . .] payout mechanism to all participants in advance”. It remains unclear whether the term “in advance” means that participants received this information several days or minutes before the experiment. Actually, the point of time of receiving this information is crucial for participants’ behavior in experiments since subjects who find out about receiving payment for their participation not before arriving at the experiment are less risk averse compared to individuals who are told one to five days before (Arkes Hal et al., 1994; Jaspersen, 2016). Another drawback may be that initial endowments varied between the two experiments. For the Philippine sample, the endowment was considerably above participants’ average daily income while it was approximately 70% of the daily income of the German students (Biener et al., 2018). To stress the impact of such payments on behavior, consider that Ackert et al. (2006) found evidence in an experiment where they varied initial endowments that traders who obtained a higher initial endowment at the start were more risk-taking resulting in higher bids.

4. Conclusion

The purpose of this paper was to examine the circumstances under which moral hazard is likely to occur and how this problem could be mitigated or eliminated. To summarize, moral hazard had a significant impact on service provision in markets for credence goods. It was found that moral hazard on the demand side, in addition to personal financial incentives for service providers to behave fraudulently, led to second-degree moral hazard. When taxi drivers were aware of the moral hazard problem between the passenger and his employer, they were more likely to overcharge (Balafoutas et al., 2017). This is in line with Kerschbamer et al. (2016) findings. However, Lu (2014) showed that the elimination of such personal incentives caused non-existence of second-degree moral hazard: Physicians who expected to receive a fraction of patients’ drug expenditures wrote 43% more expensive prescriptions to insured while doctors without financial incentives did not respond to the insurance status. Moreover, Huck et al. (2016) findings revealed the powerful effects of competition between experts to mitigate second-degree moral hazard. Indeed, insurance induced moral hazard on both market sides, but competition outweighed some of these effects. On the one hand, the consulting rate was in fact significantly higher with competition, but on the other, overtreatment was moderated. As already explained, this experiment also showed evidence of first-degree moral hazard in a health insurance context. In contrast, moral hazard was found to be less of a problem in natural disaster insurance markets with low probabilities of damages (Mol and Botzen, 2018). Insurance premium discounts and higher expected damages (i.e., higher probabilities of loss and higher deductibles) increased homeowners’ investments in damage reduction measures (Mol and Botzen, 2018). As expected, moral hazard was present when team members were paid according to their joint output. Average production levels were significantly lower and Internet usage was significantly higher under team incentives than under individual incentives. By introducing visible peer monitoring, subjects’ performance under team incentives was as high as under individ-

[49] Biener et al. (2018, p. 244)

ual incentives supporting the authors' expectation that peer pressure eliminates the problem of moral hazard in teams (Corgnet et al., 2013). Also studying joint liability group contracts, but in an insurance context, Biener et al. (2018) stressed the role of pro-social preferences to alleviate moral hazard. The higher subjects' insurance coverage was, the more decreased mean effort to self-protect against losses, but when insuring groups of two, self-protection increased because subjects were motivated by pro-social concerns.

Overall, the results suggest moral hazard to be an important problem in many markets. From the considered experimental studies, circumstances for its occurrence and mitigation measures were derived (see Appendix 10 for an overview). However, as Cohen and Siegelman (2010) proposed for proceeding with the research on adverse selection, one should identify the circumstances under which moral hazard emerges – as it was done in this paper – instead of aiming at a once-and-for-all conclusion on its existence.

In the context of second-degree moral hazard, some general research questions are still unanswered. First, little attention has been paid on how different insurance schemes such as co-payments affect sellers' provision behavior (Kerschbamer and Sutter, 2017). Second, even though insurance companies often reimburse costs for services only if they were provided by their contract partners, Kerschbamer and Sutter (2017) are not aware of studies investigating whether insurance companies' partners actually behave less fraudulently. Third, although Lu (2014) found out that experts who did not profit from overtreatment or overcharging did not prescribe insured and uninsured individuals differently, the relationship between personal incentives and provision behavior and the impact of different incentive schemes for expert sellers have not been analyzed in detail (Kerschbamer and Sutter, 2017). Fourth, the experiment with competition only implemented free choice of physicians, but the effect of price-competition may be of importance as well. Fifth, in terms of the presence of the Digital Age, one topic for future research could also be the impact of digital technologies as for example platforms like "Uber" and "sanego" on mitigating moral hazard. Such platforms may allow for decreasing informational asymmetries between sellers and customers in credence goods markets because of customers' possibilities to rate their experiences with a specific physician or to real-time monitor their "Uber" driver. And, since the previous point is in a way related to reputational incentives it may also be important to study how those affect service provision behavior.

With regard to first-degree moral hazard, future research could address the behavior of actual homeowners, instead of students, in regions that are prone to flood risk in order to increase representativeness. An additional point could be the investigation of the effect of financial incentives in natural disaster insurance markets where insurance is not mandatory. Cohen and Siegelman (2010) write that the existence of moral hazard is an important reason for using deductibles in insurance contracts. However, in the experiment by Mol and Botzen (2018), deductibles had an influence on ex ante damage reduction, but this effect was rather small which raises the question of the usefulness of high deductibles in such markets pointing out a further research topic. Furthermore, since behavior in group insurance schemes may depend on the group size it would be interesting to study the effect of different group sizes on moral hazard and, especially, whether moral hazard can also be mitigated with an increasing group size where free riding may be more likely to remain undetected and pro-sociality may decrease. Another academic void is the existence of ex post moral hazard in group insurance schemes. Finally, the finding that strategic motives can further improve subjects' effort provision in non-anonymous groups could have been driven by cultural factors since strong groups were only studied in the Philippine sample (Biener et al., 2018). Therefore, it would be interesting for follow-up research to investigate whether this effect is also robust across cultures and social classes.

References

Ackert, L. F., Charupat, N., Church, B. K., and Deaves, R. An experimental examination of the house money effect in a multi-period setting. *Experimental Economics*, 9(1):pp. 5–16, 2006.

Arkes Hal, R., Joyner Cynthia, A., Pezzo Mark, V., Gradwohl, N. J., Karen, S.-J., and Eric, S. The psychology of windfall gains. *Organizational Behavior and Human Decision Processes*, 59(3):pp. 331–347, 1994.

Arnott, R. and Stiglitz, J. E. Moral hazard and nonmarket institutions: Dysfunctional crowding out of peer monitoring? *The American Economic Review*, 81(1):pp. 179–190, 1991.

Arrow, K. Uncertainty and the welfare economics of medical care. *The American Economic Review*, 53(5):pp. 941–973, 1963.

Arrow, K. Agency and the market. In: K.J. Arrow and M.D. Intriligator, Publ. 1986. Handbook of Mathematical Economics. Vol. 3. [e-book] Amsterdam et al.: North-Holland Pub. Co. , 1986. URL http://www.ub.uni-muenchen.de/. accessed on 13.04.2019.

Balafoutas, L., Kerschbamer, R., and Sutter, M. Second-degree moral hazard in a real-world credence goods market. *The Economic Journal*, 127(599): pp. 1–18, 2017.

Berger, L. A. and Hershey, J. C. Moral hazard, risk seeking, and free riding. *Journal of Risk and Uncertainty*, 9(2):pp. 173–186, 1994.

Biener, C., Eling, M., Landmann, A., and Pradhan, S. Can group incentives alleviate moral hazard? The role of pro-social preferences. *European Economic Review*, 101:pp. 230–249, 2018.

Camerer, C. F. and Hogarth, R. M. The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, 19(1-3):pp. 7–42, 1999.

Cohen, A. and Siegelman, P. Testing for adverse selection in insurance markets. *Journal of Risk and Insurance*, 77(1):pp. 39–84, 2010.

Corgnet, B., Hérnan-Gonzalez, R., and Rassenti. Peer pressure and moral hazard in teams: Experimental evidence. Working Paper, No. 13-01, Orange, CA, Economic Science Institute at Chapman University, 2013.

Di Mauro, C. Ex ante and ex post moral hazard in compensation for income losses: results from an experiment. *The Journal of Socio-Economics*, 31(3):pp. 253–271, 2002.

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

Ehrlich, I. and Becker, G. S. Market insurance, self-insurance, and self-protection. *Journal of Political Economy*, 80(4):pp. 623–648, 1972.

Einav, L. and Finkelstein, A. Moral hazard in health insurance: What we know and how we know it. *Journal of the European Economic Association*, 16(4):pp. 957–982, 2018.

Etchart-Vincent, N. and l'Haridon, O. Monetary incentives in the loss domain and behavior toward risk: An experimental comparison of three reward schemes including real losses. *Journal of Risk and Uncertainty*, 42(1):pp. 61–83, 2011.

Finkelstein, A. and McGarry, K. Multiple dimensions of private information: evidence from the long-term care insurance market. *American Economic Review*, 96(4):pp. 938–958, 2006.

Harbaugh, W. T., Krause, K., and Vesterlund, L. The fourfold pattern of risk attitudes in choice and pricing tasks. *The Economic Journal*, 120(545): pp. 595–611, 2010.

Holmström, B. Moral hazard and observability. *The Bell Journal of Economics*, 10(1):pp. 74–91, 1979.

Holmstrom, B. Moral hazard in teams. *The Bell Journal of Economics*, 13(2): pp. 324–340, 1982.

Huck, S., Lünser, G., Spitzer, F., and Tyran, J.-R. Medical insurance and free choice of physician shape patient overtreatment: A laboratory experiment. *Journal of Economic Behavior & Organization*, 131(B):pp. 78–105, 2016.

Jaspersen, J. G. Hypothetical surveys and experimental studies of insurance demand: A review. *Journal of Risk and Insurance*, 83(1):pp. 217–255, 2016.

Karten, W., Nell, M., Richter, A., and Schiller, J. Risiko und Versicherungstechnik: Eine ökonomische Einführung. [e-book] Wiesbaden: Springer Gabler. Under: LMU Universitätsbibliothek, 2018. URL http://www.ub.uni-muenchen.de/. accessed on 13.04.2019.

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

Kerschbamer, R., Neururer, D., and Sutter, M. Insurance coverage of customers induces dishonesty of sellers in markets for credence goods. *Proceedings of the National Academy of Sciences*, 113(27):pp. 7454–7458, 2016.

Levitt, S. D. and List, J. A. What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21(2):pp. 153–174, 2007.

List, J. A. and Reiley, D. *Field Experiments. In: S.N. Durlauf and L.E. Blume, pp. 2117–2120*. The new Palgrave dictionary of economics. 2nd edition. London: Palgrave Macmillan, 2008.

List, J. A. Field experiments: A bridge between lab and naturally occurring data. *The B.E. Journal of Economic Analysis & Policy*, 6(2):pp. 1–45., 2006.

Lu, F. Insurance coverage and agency problems in doctor prescriptions: evidence from a field experiment in China. *Journal of Development Economics*, 106:pp. 156–167, 2014.

Mas, A. and Moretti, E. Peers at work. *American Economic Review*, 99(1):pp. 112–45, 2009.

Mol, J. and Botzen, J., W.J.W. and Blasch. Risk reduction in natural disaster insurance: Experimental evidence on moral hazard and financial incentives. Working Paper, No. 2018-10, University of Pennsylvania, Philadelphia, 2018.

Nell, M. Das moralische Risiko und seine Erscheinungsformen. In: Frankfurter Vorträge zum Versicherungswesen. Volume 29. Karlsruhe: Verlag Versicherungswirtschaft, pp. 7-33, 1998.

Pauly, M. V. The economics of moral hazard: Comment. *The American Economic Review*, 58(3):pp. 531–537, 1968.

Richter, A., Schiller, J., and Schlesinger, H. Behavioral insurance: Theory and experiments. *Journal of Risk and Uncertainty*, 48(2):pp. 85–96, 2014.

Salary.com. Articles. [online], 2018. URL https://www.salary.com/articles/why-how-your-employees-are-wasting-time-at-work/. accessed on 21.04.2019.

Smith, V. L. Microeconomic systems as an experimental science. *The American Economic Review*, 72(5):pp. 923–955, 1982.

Teo, T., Lim, V., and Lai, R. Intrinsic and extrinsic motivation in Internet usage. *The International Journal of Management Science*, 27:pp. 25–37, 1999.

Viaene, S. and Dedene, G. Insurance fraud: Issues and challenges. *The Geneva Papers on Risk and Insurance*, 29(2):pp. 313–333, 2004.

Yip, W. and Hsiao, W. C. The Chinese health system at a crossroads. *Health Affairs*, 27(2):pp. 460–468., 2008.