A License to Misreport: An Insurance Experiment*

Li-Hsin Lin

McMaster University, Economics Department

Abstract

Markets with asymmetric information are often characterized by the dual challenges of adverse selection and moral hazard. However, their interaction remains underexplored given the observability constraints of the real world. This study introduces a novel experimental framework that manipulates participants' awareness about their autonomy over their reported loss to investigate whether such awareness influences their willingness to pay for insurance. My results show that explicit permission to misreport increases maximum claims by 32 percentage points, while advance knowledge of the permission raises the share willing to pay the maximum price by only about 12 p.p. Interestingly, advance knowledge of the permission increases the WTP without a concurrent increase in claims. People are willing to purchase a "license" to misreport, but they do not always use it, which insurers may leverage when verification is costly.

Keywords: Selection on Moral Hazard, FFH paradigm, Dishonesty Experiment, Falsification of Claims, Insurance Demand and Indemnity

JEL: C91, D47, D82, D84, D86, L1

1. Introduction

Do individuals act in anticipation of engaging in profitable future dishonesty? In an insurance context, anticipated falsification of claims can increase individuals' willingness to pay (WTP) for coverage. This increase in WTP constitutes adverse selection on moral hazard. Consider the following scenario. Suppose your flight is severely delayed and you have travel insurance to cover

*I would like to thank the McMaster University Department of Economics for their invaluable feedback and support. I also thank the Mestelman & Muller McMaster Decision Science Laboratory (McDSL) for both financial support and helpful discussions. I benefited from helpful suggestions from participating 59th Annual Meetings of the Canadian Economics Association and 2025 Huebner-ARIA Doctoral Colloquium. Special thanks to Bradley Ruffle for support in every stage, and Katherine Cuff, Jonathan Zhang, Lanny Zrill, and Glenn Harrison for their insights and discussions. Any errors remain my own. The experiments were pre-registered.

necessary expenses. You shop at the airport and spend $150. When filing the claim, you discover that you can claim up to $250. Compared to insurance that require receipts, would you be willing to pay more for the policy that does not? This scenario illustrates how adverse selection on ex-post moral hazard can arise when the outcome state is unobservable to the principal. The framework applies broadly to markets with asymmetric information, where the answers to the opening question can inform optimal policy design.

Despite extensive theoretical, empirical, and experimental research on adverse selection and moral hazard – two central features of markets with asymmetric information – these phenomena have traditionally been studied independently. Einav et al. [4] studied the concept of "selection on moral hazard" using a quasi-experimental context of a firm's health insurance plan restructuring. This paper sparked interest in the interaction effects between adverse selection and moral hazard (Aron-Dine et al. [2]; Finkelstein et al. [5]; Shepard [8]). However, the unobservability inherent in asymmetric information markets poses challenges to such studies. As illustrated in the stylized scenario, actual outcomes of policyholders are unobservable to insurers; moral hazard is hence difficult to quantify. Similarly, assessing how much more individuals would be willing to pay for insurance that pays the indemnity with no questions asked (a measure of adverse selection) is also challenging. Meanwhile, disentangling unobservable types, for example, health conditions, financial literacy, or risk attitude, that are correlated with ex-post moral hazard and WTP presents a further complication, complicating empirical identification and control.

The experiment allows me to induce the loss distribution explicitly and directly measure both willingness to pay and moral hazard, thereby enabling clearer identification of their interaction. I introduce a relatively novel design in which the treatments manipulate the timing of when procedural instructions are revealed to participants. The approach resembles a "surprise" task or session that is commonly used to prevent data contamination. Here, it serves as the central experimental manipulation to test the core hypothesis. This design is informative for uncovering behavioral mechanisms. For instance, in the well-known dishonesty paradigm by Fischbacher and Föllmi-Heusi [6], participants roll a six-sided die in private and report the outcome for payment. Participants may choose to be dishonest about the outcome of a die roll; Yet researchers cannot distinguish between those who decide to lie before rolling and those who decide afterward. My study addresses this ambiguity by incorporating treatments that vary participants' awareness of the opportunity to

be dishonest before the revelation of the random outcomes.

Furthermore, many dishonesty experiments suffer from the lack of statistical power due to a substantial portion of participants' unconditional honesty. To address this, I design a "trick" that grants a form of license to all participants without each individual being explicitly aware that the license was granted to everyone. Furthermore, by design, participants report their loss prior to learning whether they are insured. This "quasi-strategy method" enables the collection of ex-post moral hazard outcomes for all participants, including those who ultimately remain uninsured. This approach enhances identification and broadens applicability. The results are expected to contribute to the literature of dishonesty experiments.

Adverse selection on moral hazard emerges with people's awareness of potential future moral hazard situations, particularly nuanced differences in insurance policies. In the current study, participants' awareness of future ex-post moral hazard opportunities, i.e., opportunities to misreport losses, is experimentally manipulated across treatments: in the *BASELINE* treatment, participants remain unaware that their incurred loss will be private information when purchasing insurance. To mitigate the moral costs associated with dishonesty, I design the *LICENSE* treatment, explicitly granting participants the license to claim the maximum loss before their insurance purchase decision. Lastly, the *BASELINE-LICENSE* treatment grants participants a license only after they have stated their willingness to pay for insurance. This enables direct comparisons of ex-post moral hazard with the *BASELINE* treatment, allowing for the assessment of how much maximum claims can increase when moral concerns are lifted. The comparison with the *LICENSE* treatment provides the estimation of the extent to which the increase in willingness to pay is driven by increased claims when participants are licensed. The combination of these treatments enables an evaluation, under heterogeneous types of honesty and premeditation, of whether participants strategically exploit information about their exposure to moral hazard environments by adjusting their willingness to pay.

The most relevant experimental study is Morrison and Ruffle [7], which introduced an induced loss distribution and measured both willingness to pay and inflated claims to examine the correlation between the magnitude of insurance premiums (or consumer surplus) and claim amounts. I adopt a similar induced loss distribution methodology to disentangle the connection between the paid premiums and ex-post moral hazard, without relying solely on the insured subgroup for identi-

fication. In contrast, the current study focuses on how the permission of claims falsification affects WTP, ex-post moral hazard, and their interaction.

To my knowledge, this is the first study to examine adverse selection on moral hazard with an explicitly controlled loss distribution. This methodological approach allows for clean identifications. The findings offer valuable causal and mechanistic insights into markets with asymmetric information.

The rest of the paper proceeds as follows. I present a model in Section 2. Section 3 details the experiment tasks, treatments, design, and procedures. Section 4 presents the preregistered hypotheses testing and results with the theoretical predictions. Section 5 concludes.

2. Theoretical Framework

I present a model of insurance purchasing and claims. The model allows me to more precisely define "moral hazard" and "adverse selection on moral hazard" to bridge the gap between the general framework and the experiment. It also helps me to show how the experimental environment and treatments manipulate the parameters in the model, allowing me to identify the treatment effect given heterogeneous types of dishonesty and premeditation.

2.1. A model of insurance purchasing and claims

I consider a two-period model that is designed to isolate and examine four potential determinants of Willingness to Pay (WTP) for insurance: risk aversion (ψ), the loss distribution ($F_l(l)$), forward-looking (D), and the degree of dishonesty (θ^{RH}). In the first period, risk-averse, expected-utility-maximizing individuals determine their WTP for insurance. Their decisions are based on their type along three dimensions: loss distribution, forward-looking, and the degree of dishonesty. Each individual forms expectations about future claims using the available information. In the second period, individuals observe their realized losses and determine their claims. I closely follow the framework of Einav et al. [4]. I begin by presenting a general form of the second period that encompasses both the specifications in Einav et al. [4] and my extensions. Then, I incorporate a behaviorally calibrated utility function drawn from Abeler et al. [1], which captures a "preference for truth-telling." I show that the dishonesty parameter from Abeler et al. [1], which captures heterogeneity in truth-telling preferences, can be interpreted similarly to the parameter in Einav

et al. [4] that governs an individual's price elasticity of demand for health care with respect to out-of-pocket expenditures. In the first period, I incorporate premeditated types. This addition contributes to the literature by incorporating uncertainty in the claims process as perceived by individuals when making insurance purchasing decisions. Furthermore, it enables heterogeneous types of premeditation (i.e., degrees of forward-looking behavior).

2.2. Second Period - Utility from Misreport

Consider the following general form of individuals' utility in the second period:

$$u(l; r, \cdot) = (r - l) - c(l; r, \cdot) + y(\cdot) \tag{1}$$

where l is the true loss, r is the reported loss and $c(l; r, \cdot)$ is the cost of misreporting, or over-utilization. The $(\cdot)$ here is reserved for variables that lead to heterogeneous misreporting costs; while $y(\cdot)$ is the residual income, including the residual income, co-payment, and the price paid for the insurance.

I deviate from Einav et al. (2013) by specifying the $c(l; r, \cdot)$ function using a calibrated utility cost of dishonesty from Abeler et al. (2019):

$$u_i(l; c(r, l), \Lambda(r), \theta^{\mathrm{RH}}) = r - c\mathbb{I}_{l \neq r} - \theta_i^{\mathrm{RH}}\Lambda(r) - l + y_i \tag{2}$$

$\Lambda(r)$ is the fraction of liars at r; c is a binary fixed cost of lying which depends on whether an individual lied. The parameter that governs the individual's type of dishonesty, the reputation for honesty, is θ^{RH}. In Abeler et al. [1], this is an individual-specific weight on reputation, drawn from a uniform distribution on $[0, \kappa^{rh}]$. The return to dishonesty is $r - l$ and the cost is $c\,\mathbb{I}_{l \neq r} + \theta_i^{\mathrm{RH}}(\Lambda(r) - \Lambda(l))$; The larger θ_i^{RH} is, the less likely the falsification of claims is and the smaller of the falsified claim is. Compared to the healthcare framework in Einav et al. [4], we can interpret θ^{RH} as the inverse of the individual's price elasticity of demand for health care with respect to its (out-of-pocket) price.

The optimal claim is given by:

$$r_i^*(l, \Lambda(r), \theta_i^{\text{RH}}) = \arg\max_{r \geq 0} u_i(r; l, \Lambda(r), \theta_i^{\text{RH}}) \tag{3}$$

Before proceeding to the first period, I propose two extensions to the theoretical framework to relax the assumptions of perfect forward-looking behavior and perfect execution. First, I introduce a parameter $A \in [0, 1]$, representing the proportion of participants who form their willingness to pay (WTP) based on anticipated utility from misreporting. The remaining share, $1 - A$, comprises individuals who demand insurance without contemplating the potential for dishonesty. This parameter allows for a richer interpretation, as perfect forward-looking behavior may depend on knowledge of the indemnity process, financial literacy, and cognitive sophistication. Second, I allow the dishonesty type parameter θ_i^{RH} to vary across periods, denoted $\theta_{i,t=1,2}^{\text{RH}}$, to capture imperfect execution. For example, an individual may purchase insurance with the intention of misreporting but ultimately submit an honest claim, or vice versa. The experimental design exploits participants' knowledge of the indemnity process and elicits both their WTP and falsified claims. This enables the identification of forward-looking behavior and temporal consistency in dishonest reporting within the insurance market.

2.3. First Period - Willingness to Pay

Expected-utility-maximizing individuals with a coefficient of relative risk aversion of ψ follow the constant relative risk aversion (CRRA) form $u(x) = \frac{x^{1-\psi}}{1-\psi}$ (if $\psi \neq 1$). With a proportion or probability A that $D = 1$ otherwise $D = 0$, the expected utility of the insured individual is:

$$\begin{aligned} v_i(F_l(\cdot), c(r, l), \Lambda(r), \theta_{i,1}^{\text{RH}}, \psi) = (1 - D) \int (u^*(l))^{1-\psi} \, dF_l(l) \\ + D \int \left(u^*(c(r, l), \Lambda(r), \theta_{i,1}^{\text{RH}})\right)^{1-\psi} \, dF_l(l) \end{aligned} \tag{4}$$

The first term is the second-period expected utility from honest reports, and the second term is the premeditated dishonest part that would affect $v_j(\cdot)$, the WTP, causing adverse selection on moral hazard. On the other hand, the expected utility of the uninsured individual is:

$$v_i'(F_l(\cdot), \psi) = \int (-l + y_i)^{1-\psi} \, dF_l(l) \tag{5}$$

So the willingness to pay for insurance is:

$$WTP_i = v_i(F_l(\cdot), c(r, l), \Lambda(r), \theta_{i,1}^{\mathrm{RH}}, \psi) - v_i'(F_l(\cdot), \psi) \tag{6}$$

Note that adverse selection on ex-post moral hazard in this model implies that $\frac{\partial \mathrm{WTP}_i}{\partial \theta_{i,2}^{\mathrm{RH}}} > 0$, which requires two key conditions beyond those in the baseline framework. First, forward-looking behavior must be present, i.e., $A \neq 0$, such that some participants form their willingness to pay based on anticipated claim behavior. Second, dishonesty types must be consistent across time: $\theta_{i,2}^{\mathrm{RH}} \propto \theta_{i,1}^{\mathrm{RH}}$. That is, ex-post dishonesty should be at least partially predictable from ex ante intentions. Conversely, two types of individuals do not contribute to this form of adverse selection: (1) those who did not consider the possibility of falsifying claims at the insurance purchase stage ($D = 0$); and (2) those who engage in dishonesty spontaneously or inconsistently, such that $\theta_{i,2}^{\mathrm{RH}} \perp\!\!\!\perp \theta_{i,1}^{\mathrm{RH}}$. For both, the derivative of WTP with respect to their eventual dishonesty type is zero: $\frac{\partial \mathrm{WTP}_i}{\partial \theta_{i,2}^{\mathrm{RH}}} = 0$. Hence, adverse selection on moral hazard in this setting is fundamentally driven by the joint presence of forward-looking decision-making and intertemporal consistency in dishonest behavior.

2.4. Experiment Identification

The design has two key advantages. First, I observe willingness to pay (WTP) directly. In empirical settings, WTP is typically inferred from discrete insurance choices; here, identification does not rely on such indirect variation. Second, I induce a single, known loss distribution common to all participants. This breaks any potential correlation between $F_l(\cdot)$ and heterogeneity in forward-looking and dishonesty types (D, θ^{RH}), and it allows me to measure ex-post moral hazard as the gap between reports and losses.

License. I exogenously shift the degree of dishonesty using a simple "license" that explicitly permits claiming the maximum loss. Providing the license *before* insurance purchase (First Period) or *after*

(Second Period) changes $(\theta_{i,1}^{\mathrm{RH}}, \theta_{i,2}^{\mathrm{RH}})$ differentially. Relative to *BASELINE* (license never given), *BASELINE–LICENSE* offers the license only in the Second Period. The procedures up to insurance purchase are identical across these two conditions in the First Period, so the WTP should remain unchanged and $\theta_{i,1}^{\mathrm{RH}}$ remains the same. By contrast, the First Period license induces a shift in the degree of dishonesty $\Delta\theta_{i,2}^{\mathrm{RH}}$, the treatment effect on ex-post moral hazard is:

$$r_i^*(BASELINE - LICENSE) - r_i^*(BASELINE) = \frac{\Delta r_i^*(l, \Lambda(r), \theta_{i,2}^{\mathrm{RH}})}{\Delta\theta_{i,2}^{\mathrm{RH}}} \tag{7}$$

Comparing *BASELINE-LICENSE* to *LICENSE* for which the license was shown to participants in the First Period, assuming the license shifts the degree of dishonesty similarly in both periods, the shift on the expectation of the degree of dishonesty and the shift on the degree of dishonesty when reporting. That on group level, $\Delta\overline{\theta_{i,1}^{\mathrm{RH}}} = \Delta\overline{\theta_{i,2}^{\mathrm{RH}}}$, and the average treatment effect on WTP is:

$$\frac{\Delta\overline{WTP}}{A\Delta\overline{\theta_{i,1}^{\mathrm{RH}}}} = \frac{\Delta\overline{WTP}}{A\Delta\overline{r_i^*(l, \Lambda(r), \theta_{i,1}^{\mathrm{RH}})}} \frac{\Delta\overline{r_i^*(l, \Lambda(r), \theta_{i,2}^{\mathrm{RH}})}}{\Delta\overline{\theta_{i,2}^{\mathrm{RH}}}} = \frac{\Delta\overline{WTP}}{A\Delta\overline{r_i^*(l, \Lambda(r), \theta_{i,1}^{\mathrm{RH}})}} ATE_{r_i^*} \tag{8}$$

Note that the existence of both treatments' differences in ex-post moral hazard and WTP here provides no evidence for adverse selection on ex-post moral hazard. We could have $\Delta\overline{\theta_{i,1}^{\mathrm{RH}}} = \Delta\overline{\theta_{i,2}^{\mathrm{RH}}}$ while $\theta_{i,2}^{\mathrm{RH}} \perp\!\!\!\perp \theta_{i,1}^{\mathrm{RH}}$ as the case that our average treatment effects is exactly the same as in another case $\theta_{i,2}^{\mathrm{RH}} = \theta_{i,1}^{\mathrm{RH}}$. We will need to compare the WTP distribution and the Reports distribution to learn about $\frac{\partial WTP_i}{\partial\theta_{i,2}^{\mathrm{RH}}}$.

3. Experiment

3.1. Tasks

A questionnaire was administered at the beginning of the experiment to collect demographic information. Following this, participants completed the Holt-Laury risk elicitation task (see Table 1), referred to as Task 1, which measures individual risk preferences. Subsequently, participants received instructions for Task 2. These instructions informed participants that they had earned \$13 from completing the questionnaire, of which \$8 is subject to a potential loss. They were then offered an opportunity to purchase insurance against the full value of this potential loss. Treatment-specific instructions were introduced at this stage to manipulate participants' awareness of their potential

exposure to an ex-post moral hazard situation. To ensure comprehension, a detailed quiz was administered prior to the insurance purchasing decision.

Participants' willingness to pay (WTP) and their insurance decisions were elicited using a Multiple Price List (MPL) format (see Table 2). The design of Task 2 was adapted from Morrison and Ruffle [7], incorporating modifications to align with the specific goals of this study. In the loss realization phase, an experimenter presented a transparent basket containing numerous opaque, identical containers, each holding \$8 minus a potential loss. Participants were informed that possible losses are \$8, \$6, \$4, \$2, \$1, or \$0, each with an equal 16.67% (1/6) probability of being selected. Participants sequentially and privately selected one container without replacement and opened it at their station to discover their assigned loss. Participants then reported this loss, or any amount they chose, on a reporting page. After the reporting phase, participants learned whether they were insured and the insurance price. The results page displayed their earnings from each component of the experiment, along with their total payment, excluding the privately observed amount in their containers.

Table 1: Holt-Laury Risk task

Decision	Option A				Option B				Your Decision	
		DOLLARS		DOLLARS		DOLLARS		DOLLARS	I choose Option A	I choose Option B
1	A 10% chance of	$5	A 90% chance of	$4	A 10% chance of	$8	A 90% chance of	$0.50	○	○
2	A 20% chance of	$5	A 80% chance of	$4	A 20% chance of	$8	A 80% chance of	$0.50	○	○
3	A 30% chance of	$5	A 70% chance of	$4	A 30% chance of	$8	A 70% chance of	$0.50	○	○
4	A 40% chance of	$5	A 60% chance of	$4	A 40% chance of	$8	A 60% chance of	$0.50	○	○
5	A 50% chance of	$5	A 50% chance of	$4	A 50% chance of	$8	A 50% chance of	$0.50	○	○
6	A 60% chance of	$5	A 40% chance of	$4	A 60% chance of	$8	A 40% chance of	$0.50	○	○
7	A 70% chance of	$5	A 30% chance of	$4	A 70% chance of	$8	A 30% chance of	$0.50	○	○
8	A 80% chance of	$5	A 20% chance of	$4	A 80% chance of	$8	A 20% chance of	$0.50	○	○
9	A 90% chance of	$5	A 10% chance of	$4	A 90% chance of	$8	A 10% chance of	$0.50	○	○
10	A 100% chance of	$5	A 0% chance of	$4	A 100% chance of	$8	A 0% chance of	$0.50	○	○

Note: Participants were asked to choose options A or B for each row to assess their risk aversion.

Table 2: Multiple Price List for insurance

Price of Insurance Contract	Yes, I am willing to pay this price	No, I am not willing to pay this price
$0	○	○
$1	○	○
$2	○	○
$3	○	○
$4	○	○
$5	○	○
$6	○	○
$7	○	○

3.2. Treatments: Timing of License

I designed treatments that manipulated information regarding the potential for dishonesty. In the *BASELINE* (**BB**), the instructions merely informed participants that $8 of their questionnaire earnings was subject to a random loss. The *BASELINE-LICENSE* (**BL**) and **LICENSE** (**LL**) incorporated a "condition" granting all participants a license to report any loss amount up to $8. This was introduced through the following instruction:

> "Additionally, if the sum of the digits in your birthdate (written in MMDD format) is less than 21, you are allowed to report any loss up to 8, no matter the actual loss amount. For example, if your birthday is August 19 (written as 0819), the digits 0 + 8 + 1 + 9 add up to 18. Since 18 is less than 21, you may report any loss up to 8."

The maximum MMDD combination is September 29 (0929), for which the digits sum to 20, still less than 21. Therefore, all participants effectively receive this license. The three treatments differ according to whether this reporting license is never provided (**BB**), provided before the loss reporting stage (**BL**) or provided both before the Multiple Price List (MPL) used to elicit the participant's WTP for insurace and before the loss-reporting stage (**LL**). When the license was introduced before the MPL stage, participants had the opportunity to condition their willingness to pay on anticipated dishonesty, thus enabling the identification of adverse selection on ex-post moral hazard. When the license was introduced only before the reporting stage, after the insurance purchasing decision had already been made, the experiment isolates pure moral hazard effects. This design allows for a clean comparison of behaviors under different informational and temporal regimes, enabling the disentanglement of premeditated dishonesty from opportunistic moral hazard.

3.3. Procedures

The experiment was conducted at the Mestelman & Muller McMaster Decision Science Laboratory (McDSL), which is equipped with 22 computerized partitioned stations. The experimental tasks were programmed using oTree (Chen et al. [3]), with the exception of the physical loss generation process in opaque containers, which was manually implemented as described previously. Participants were randomly assigned to individual cubicles after reading the letter of information and signing the consent form. Upon being seated, they were welcomed with an on-screen message

and a comprehensive set of instructions. To ensure clarity and consistency in comprehension, the instructions were also provided in printed form and played as an audio recording through each participant's headset.

The total payoff for an insured participant included the following components: a $5 fixed payment for completing the questionnaire, a payment from the Holt-Laury Risk task, up to $1 based on performance on the knowledge-testing quiz (depending on the number of correct answers on the first attempt), the claim amount minus the insurance premium (if insured), and the value retrieved from the loss container ($8 that they earned from the questionnaire, minus the actual randomly determined loss). The last payment component was delivered during the experiment using the physical container method, while all other payments were made in cash upon completion of the session. For an uninsured participant, the total payoff comprised the questionnaire payment, Holt-Laury task earnings, quiz performance bonus, and $8 minus the actual loss drawn from the container.

3.4. Participants

To recruit participants for the experiment, the SONA participant pool management system was utilized, drawing from the existing pool maintained by the McDSL. The sample size is 162 participants, with 51 participants in **BB**, 58 participants in **BL**, and 53 participants in **LL**, consisting of three between-subjects treatments. The mean of payments was $20.79 CAD for a 45-minute-to-one-hour session, which included a $5 CAD show-up fee.

4. Results

In this study, I used the license, exogenously and at different points in time, to reduce the dishonesty cost, allowing me to test if the dishonesty is anticipated (selection) and if the anticipation would result in ex-post moral hazard. My preregistered hypotheses are based on the assumption that the license leads to more dishonesty, the license increases WTP, and the high WTP group displays more dishonesty.

4.1. Preregistered Hypotheses Testing

Report - Ex-Post Moral Hazard.

Since the loss distributions were identical across treatments, any variation in the reported losses

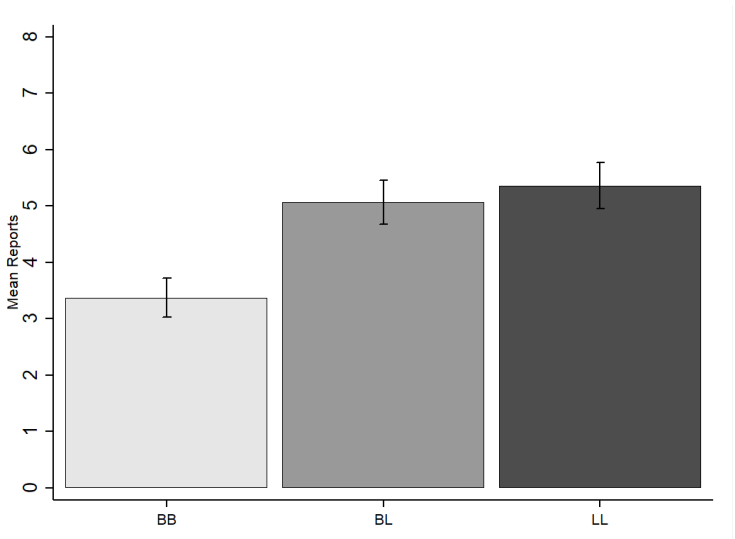
can be interpreted as a difference in ex-post moral hazard, specifically, the deviation between the reported and the true losses.

If we exclude the moral hazard that arises due to earlier awareness of the license, the variation in moral hazard must be attributed to the change in dishonesty cost by license at the reporting stage:

H1 (Preregistered): **LL** = **BL** > **BB**

Notably, the observed differences in reporting behavior are driven by comparisons with participants who received the *BASELINE* version of the instructions before loss reporting. A significant difference suggests that licensing can affect behavior even among those who would otherwise act honestly in **BB**.

H1 is verified. Mean reports of **LL**, **BL**, and **BB** are 5.36, 5.07, and 3.37 respectively (see Figure 1). The difference between **LL** and **BL** is insignificant ($t-test,\ p=.31$); while both reports in **LL** and **BL** are significantly larger than in **BB** ($t-test,\ p<.001;\ p<.001$). As a benchmark, note that the mean of true loss is 3.50.



Notes: The possible losses are \$8, \$6, \$4, \$2, \$1, or \$0, with 3.50 as the mean. The true realized mean loss by treatment **LL**, **BL**, and **BB** are 3.27, 3.38, and 3.55 respectively. On average, participants in **LL** and **BL** over-reported by 1.75.

Figure 1: Mean report by treatments

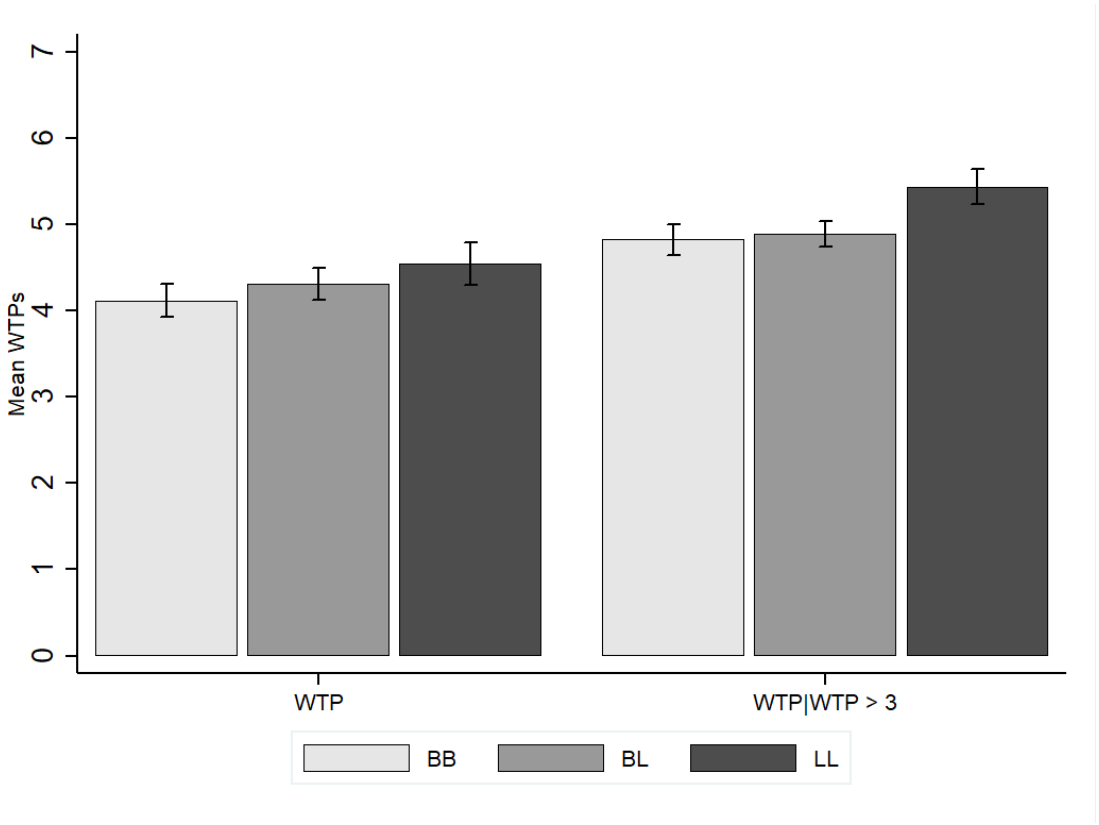
Result 1: The license in the report stage increased participants' falsification of claims; the license in the insurance purchasing stage did not further increase the falsification of claims.

Willingness To Pay - Anticipated Ex-Post Moral Hazard.

Honest participants ought to value insurance according to their risk premium (equivalent to the case $D = 0$ in the theoretical framework section), whereas participants who anticipate the potential to falsify their reports ought to value insurance at their certainty equivalent given their intended misreports at different losses (the case $D = 1$). The difference in valuation should be driven by the participants' awareness of the upcoming moral hazard opportunity, which was shaped by the license instruction they received.

H2 (Preregistered): **LL** > **BL** = **BB**

H2 is partially verified. Mean WTP of **LL**, **BL**, and **BB** are 4.55, 4.31, and 4.12, respectively (see Figure 2). Only WTP of **LL** are marginally significantly larger than **BB** ($t-test,\ p < .10$), with no significant differences between **LL** and **BL** nor **BL** and **BB** ($t-test,\ p = .21;\ p = .23$). However, in theory, for a risk-averse individual who is the target of the insurance, they should have a positive risk premium or be willing to pay more than mean of the loss distribution, i.e., WTP greater than 3.5. Conditional on WTP greater than or equal to 4, WTP of **LL** are significantly greater than those of **BL** and **BB** ($t-test,\ p < .03;\ p < .03$), with mean of WTP in **LL**, **BL**, and **BB** are of 5.43, $n = 37$, 4.89, $n = 44$, and 4.82, $n = 34$, respectively.



Notes: WTP ranged from 0 to 7. The left panel is the mean WTP, and the right panel is the mean WTP excluding participants with WTP below 4 since they prefer the risk over the insurance.

Figure 2: Mean WTP by treatments

Result 2: Conditional on participants willing to pay more than the mean of the loss distribution, when the license to misreport is granted before the insurance-purchase decision, the willingness to pay for insurance increases significantly.

Report - Adverse Selection on Ex-Post Moral Hazard.

Do premeditated intentions to be dishonest translate into falsified claims? The experimental design allows me to measure ex-post moral hazard unconditionally, whether participants are insured or not. This question can be addressed by comparing the level of dishonesty conditional on the subset of participants with higher WTP and lower WTP, as the high level dishonest type would self-select into higher WTP from lower WTP. Their claim reports are expected to differ according to whether

the license was provided before the insurance-purchasing decision:

H3 (Preregistered-modified): **LL** > **BL** > **BB** (conditional on High-WTP)

H3.1 (Additional): **BL** > **LL** > **BB** (conditional on Low-WTP)

H3 is partially verified. There are four insurance premiums above the mean of the loss distribution: 4, 5, 6, *and* 7. I take 6 and 7 as my criterion for High-WTP. Conditional on High-WTP, mean reports of **LL**, **BL**, and **BB** are 5.82 ($n = 17$), 4.78 ($n = 9$), and 4.29 ($n = 7$) respectively. The difference in reports between **LL**, **BL** and **BB** are not significantly different from zero at conventional levels ($t - test,\ p = .21;\ p = .12;\ p = .38$). On the other hand, conditional on WTP of 4 or 5 (Low-WTP) mean reports of **LL**, **BL**, and **BB** are 4.40 ($n = 20$) 5.51 ($n = 35$), and 3.11 ($n = 27$), respectively. Reports of **LL** are now marginally significant lower than those of **BL** ($t - test,\ p < .1$) due to selection, while both reports of **LL** and **BL** are significantly larger than those of **BB** ($t - test,\ p < .1;\ p < .001$). However, the interaction term of High-WTP/Low-WTP and **LL**/**BL** using mixed effects is insignificant ($t - test,\ p = .12$).

WTP and Report - Positive Correlation.

Alternatively, I tested for a positive correlation between WTP and reports across treatment groups.

H4 (Preregistered-modified): The positive correlation between WTP and reports by treatment is expected to be: **LL** > **BB** > **BL**

The rationale is as follows. In **BB**, participants anticipate some degree of ex-post moral hazard, which induces a weak positive correlation between WTP and subsequent reports. Introducing the license only at the claims stage breaks this link, because the realized policy at reporting differs from the policy participants expected at purchase. By contrast, when the license is disclosed already at purchase (**LL**), the correlation is reintroduced via adverse selection on ex-post moral hazard.

Support for H4 is minimal. Conditional on WTP of 4, the correlation between WTP and reports of **LL**, **BL**, and **BB** are $r = 0.23,\ p = .17$, $r = -0.07,\ p = .64$, and $r = 0.16,\ p = .36$, respectively. Excluding **BL**, conditional WTP of 4, the correlation between WTP and reports of **LL** and **BB** combined is $r = 0.27,\ p < .03$.

Result 3: Despite suggestive patterns, we find no clear evidence of adverse selection

on ex-post moral hazard.

4.2. Anticipation of Dishonesty vs. Spontaneous Dishonesty

Conversely, if one interprets **Result 3** as a null (or negative) finding, then taken together **Results 1-3** motivate extending the Einav et al. [4] framework to allow (i) heterogeneity in anticipatory dishonesty: not everyone adjusts WTP in anticipation of misreporting, and (ii) execution consistency at the claims stage: individuals may become more or less dishonest when filing. Empirically, comparing reported losses across the three treatments indicates that the Second Period license shifts the (implicit) cost of dishonesty in a similar way across conditions, whereas the First Period license does not affect the reports. By contrast, comparing WTP across treatments shows that WTP responds to anticipated dishonesty, which is affected by the First Period license.

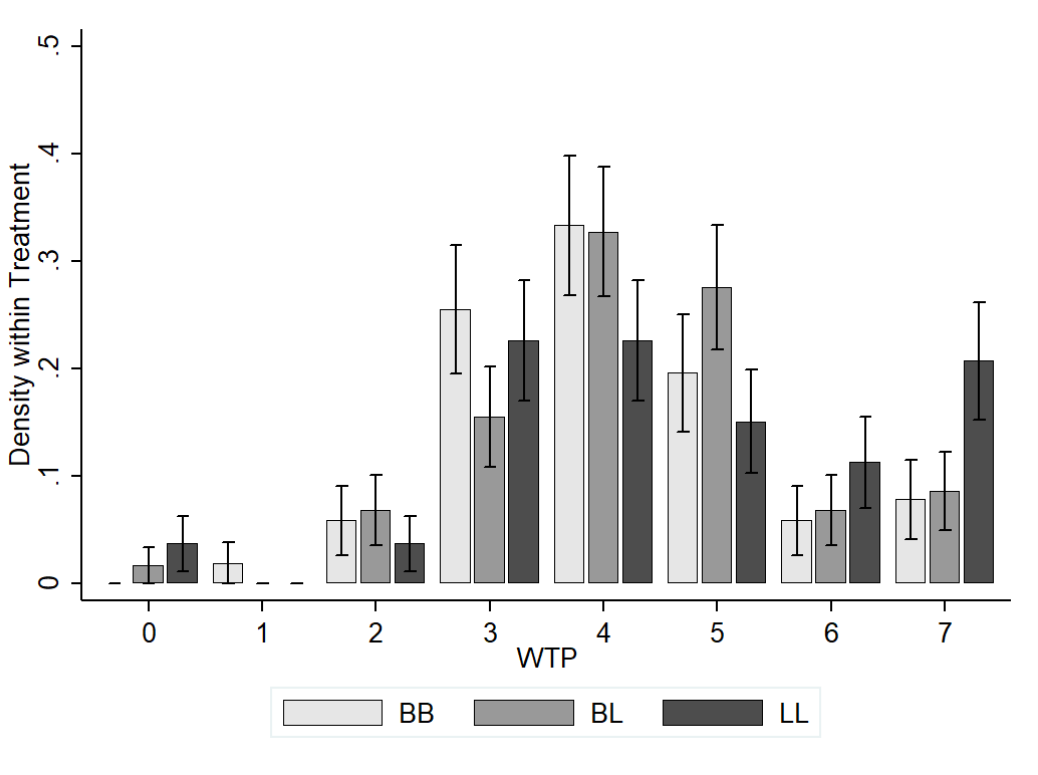
A key contribution of the paper is to demonstrate that **Result 1** and **Result 2** by themselves do not imply adverse selection on moral hazard. Only when accounting for both anticipatory adjustments (affecting WTP) and claims-stage execution (affecting reports) does the full mechanism become apparent.

WTP distribution. Comparing **BL** to **LL**, the shares at each WTP are: 33% vs. 23% at WTP = 4 (*proportion test*, $p = .12$); 28% vs. 15% at WTP = 5 (*proportion test*, $p < .10$); 7% vs. 11% at WTP = 6 (*proportion test*, $p = .21$); and 9% vs. 21% at WTP = 7 (*proportion test*, $p < .05$) (see Figure 3). Thus, showing the license at purchase shifts mass away from WTP of 5 and toward WTP of 7.

Reports by WTP. The mean reported losses (by WTP) in **BL** vs. **LL** are: 5.47 vs. 3.83 at WTP = 4 ($t-test$, $p < .10$); 5.56 vs. 5.25 at WTP = 5 ($t-test$, $p = .40$); 4.25 vs. 6.67 at WTP = 6 ($t-test$, $p < .10$); and 5.20 vs. 5.36 at WTP = 7 ($t-test$, $p = .47$) (see Figure 4). Hence, the license at purchase is associated with fewer falsifications at WTP of 4 and more at WTP of 6.

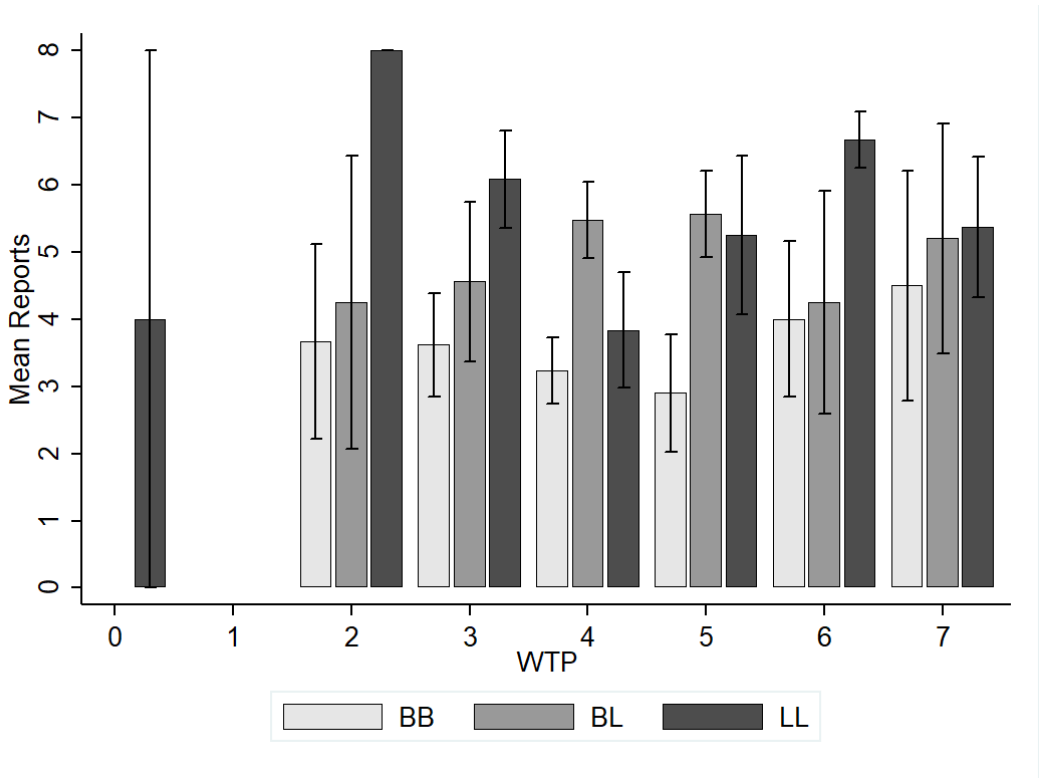
Interpretation. Aligning the WTP shifts with the significant report differences suggests that participants who increase WTP in response to the purchase-stage license falsify "as planned" only up to a limited proportion.

Hot-spot Robustness. Using reports of 6 or 8 as dishonesty "hot-spots," I find: conditional on Low-WTP, the hot-spot share is 60% in **BL** vs. 45% in **LL** (*proportion test*, $p = .15$); conditional on High-WTP, it is 56% in **BL** vs. 76% in **LL** (*proportion test*, $p = .15$). A mixed-effects interaction of High-WTP/Low-WTP with **LL**/**BL** is not statistically significant ($p = .13$).



Notes: At Low-WTP, densities of **LL** are lower than both **BB** and **BL**; At High-WTP, densities of **LL** are higher than both **BB** and **BL**. The WTP distribution right shifted when the license was given at the insurance purchasing stage.

Figure 3: WTP distribution by treatments



Notes: The main reduction happened at WTP of 4 and main increase happened at WTP of 6. However, the main increased distribution was at WTP of 7.

Figure 4: Mean report by treatments and WTP

Forward-looking ratio. Conditional on WTP ≥ 4, the share of dishonesty hot-spot reports (6 or 8) rises from **BB** to **BL** by 35.6 percentage points (from 23.5%, $n = 34$, to 59.1%, $n = 44$; *proportion test*, $p < .001$). Over the comparison between **BL** to **LL**, the share of participants with High WTP increases by 25.4 p.p. (from 20.5%, $n = 44$, to 45.9%, $n = 37$; *proportion test*, $p < .01$). Interpreting these movements in the model, the fraction of participants who anticipated potential dishonesty (i.e., forward-looking) is $A = \frac{25.4}{35.6} = 71.35\%$.

Claims-stage consistency. In **BL**, 5 of 58 participants report WTP = 7; in **LL**, 11 of 53 do so (more than double), an increase of 12.2 p.p. (from 8.6% to 20.8%; *proportion test*, $p < .05$). Yet among these WTP = 7 participants, hot-spot reporting is 3/5 in **BL** and 7/11 in **LL**, a rise of only 3.6 p.p. (from 60.0%, $n = 5$, to 63.6%, $n = 11$; *proportion test*, $p = .45$). The near-constancy of the hot-spot share across treatments indicates a weak translation from higher WTP to realized falsification at the top of the WTP distribution.

Policy implication. One might infer from Einav et al. [4] that if claim falsification (ex-post moral hazard) — or, here, the license — is unavoidable, it should be obscured at purchase. The behavioral evidence here suggests the opposite: greater transparency can raise uptake (via higher WTP) without proportionally increasing realized falsification, potentially benefiting even the insurer. In this experimental setting, with randomized insurance prices and conditioning on WTP > 3, expected payouts are slightly higher in **BL** (estimated \$22.19) than in **LL** (estimated \$21.95).

Result 4: License-induced misreporting is not always accompanied by higher willingness to pay, indicating partial forward-looking.

Result 5: License-induced increases in WTP are not always accompanied by falsified claims, indicating a level of inconsistency of dishonesty.

Result 6: Insurers may benefit from offering the license at the purchase stage if the resulting increase in uptake offsets the cost arising from selection on ex-post moral hazard.

5. Conclusion

Many insurance markets, because of inherent asymmetric information, contend with both adverse selection and moral hazard. Empirically, these problems are difficult to study: observability of willingness to pay (WTP) and realized losses is limited, and loss distributions are typically endogenous, let alone analyzing their interaction. My experiment directly measures participants' WTP and induces a common, known loss distribution that is independent of participants' characteristics. Methodologically, I introduce a "license" procedure that exogenously shifts the cost of dishonesty without participants realizing that it is a treatment-wide manipulation. I further minimize confounds by informing participants whether they are insured and the price paid, thereby avoiding effects driven by realized consumer surplus or premium salience, while still collecting claims from all participants, without selection. Finally, by eliciting indications of dishonesty at two distinct

stages, I show that not all false reports are premeditated: some participants who appear "interested" in dishonesty ex ante subsequently change their minds at the reporting stage.

The results also speak directly to the insurance literature. A prominent line of work asks whether increasing the salience of policy features can raise interest in coverage and mitigate underinsurance. This experiment provides clean evidence that an additional, concrete piece of information about the indemnity process, specifically, what claimants can do, shifts willingness to pay. Moreover, many instruments designed to limit moral hazard rely on frictions at the claims stage ("nitpicking"), which are effectively the opposite of a no-questions-asked license. Such frictions can exacerbate underinsurance and impose administrative costs on insurers. By contrast, committing at purchase to "no nitpicking" can increase WTP and policy uptake, and the selection on moral hazard need not raise realized ex-post moral hazard. Behaviorally, I show that anticipated ex-post moral hazard affects WTP, but the induced increase in WTP does not necessarily translate into higher realized moral hazard at the claims stage. The finding sheds light on the literature considering selection on moral hazard as a mechanism behind positive correlations between coverage and claims/utilization. The findings imply that, in environments where indemnity verification is difficult, insurers should consider a guaranteed no-questions-asked (or limited-friction) policy. Doing so both addresses underinsurance and resolve the ambiguity of indemnity as the folklore claim that "you only find out whether you're insured when you make a claim."

References

[1] Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, 87(4):1115–1153.

[2] Aron-Dine, A., Einav, L., Finkelstein, A., and Cullen, M. (2015). Moral hazard in health insurance: Do dynamic incentives matter? *Review of Economics and Statistics*, 97(4):725–741.

[3] Chen, D. L., Schonger, M., and Wickens, C. (2016). oTree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

[4] Einav, L., Finkelstein, A., Ryan, S. P., Schrimpf, P., and Cullen, M. R. (2013). Selection on moral hazard in health insurance. *American Economic Review*, 103(1):178–219.

[5] Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., and Group, O. H. S. (2012). The Oregon health insurance experiment: Evidence from the first year. *Quarterly Journal of Economics*, 127(3):1057–1106.

[6] Fischbacher, U. and Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, 11(3):525–547.

[7] Morrison, W. G. and Ruffle, B. J. (2025). Do higher insurance premiums provoke larger reported losses? An experimental study. *Journal of Risk and Insurance*, 92(1):203–226.

[8] Shepard, M. (2022). Hospital network competition and adverse selection: Evidence from the massachusetts health insurance exchange. *American Economic Review*, 112(2):578–615.