Effects of Incentives and Education on Financial Choices: An Experiment*

Han Huynh†

December 2, 2019

Abstract

Mounting evidence reveals a puzzle in consumer finance: in high-stakes financial decisions, people leave a substantial amount of money on the table, even when financial education is available. The ubiquity of financial choices makes understanding the effects of incentives and education on mistakes crucial. This project experimentally examines the impact of changes in incentives and educational availability on incentivized but hypothetical healthcare choices using Amazon Mechanical Turk. We find that increasing incentives are ineffective in increasing decision-making effort, even when these changes are made clear and salient to the subjects. Yet, surprisingly, despite this lack of effort response, subjects' choices improve when incentives are high. This result highlights an under-appreciated channel of incentives: when stakes become larger, often, the problems become simpler too. We next investigate the effect of available education. Overall, education leads to an increase in decision-making effort and an improvement in choice quality. However, this average effect masks significant heterogeneity across incentive treatments. Subjects are willing to put in the educational effort when either the problems are hard or mistakes are highly costly, but the return of the educational effort is zero for hard problems and positive for easy ones. Thus, only when stakes are high and the problem is easy does education have an effect. These findings suggest that people can be encouraged to get education for high-stakes decisions, and policy-makers have a role in simplifying problems to translate the extra effort into better choices.

*I especially thank Mark Dean for his unwavering support and his incomparable comments and suggestions. I also thank Jack Willis and Alessandra Casella for their invaluable feedback. I am grateful to Eric Johnson, Navin Kartik, Yeon-koo Chee, and the participants of the Cognitive and Decision Lab, the Experimental Lunch, and the Micro-economic Theory Colloquium at Columbia University for their suggestions at various stages of the project. I thank Leo Goldman for his excellent assistance with Amazon Mechanical Turk. I gratefully acknowledge financial support from the National Science Foundation (grant SES-1919483), the Columbia Experimental Laboratory for Social Sciences, the Microeconomic Colloquium, and the Microeconomic Theory Initiative.

†Department of Economics, Columbia University

1 Introduction

Research in household finance has shown that consumers make financially sub-optimal choices, thus, leaving a substantial amount of money on the table in various types of financial decisions (loans in Agarwal et al. (2009), Bertrand and Morse (2011), insurance in Bhargava et al. (2015), Abaluck and Gruber (2011), and investment in Beshears et al. (2011)). One obvious explanation is that people lack the skills to make these decisions well. Yet, the literature finds that these mistakes persist even when financial education is available. A recent meta-analysis on financial literacy and financial behaviors finds that education has surprisingly little effects on choices (Fernandes et al. (2014)). Why do consumers choose financially sub-optimal products even though mistakes are highly costly and education is freely available?

As a starting point to answering the question, this project designs an experiment to identify the effects of incentives and education on financial choices. This experiment mimics the choice of health insurance, a setting in which consumers incur significant losses and often misunderstand how the products work (Bhargava et al. (2015), Loewenstein et al. (2013)).[1] We recruit Amazon Mechanical Turk (MTurk) workers to choose insurance plans for hypothetical scenarios. Each scenario consists of deterministic health-care needs so that the total health costs of all listed plans are deterministic.[2] The objectively correct choice is then the lowest-cost plan, and subjects receive a higher payment for choosing a lower-cost plan. Within this task, we vary the stakes and access to education. The variations allow us to test whether incentives and education matter for choices, the mechanisms by which they do so, and the effects of their interaction.

To vary the stakes, we design low- and high-incentive treatments. Each subject is randomized into one incentive level, which corresponds to a high or low cost of an average mistake. We make incentives higher by changing the premiums such that the variance of the total health costs increases. Because this increases the difference between the best plan and a randomly chosen plan is higher, mistakes become more costly.

We first look at the impact of incentives on how much effort subjects put into choosing a plan. When people make bad choices even with high stakes, there are two possible explanations. Either people do not increase effort with high stakes, or the extra effort is in vain. Using time spent on the task as a proxy for effort, we find evidence for the former hypothesis: subjects do not increase effort when incentives are higher.[3] There is no effect of high stakes on time, although the best plan in the high-incentive treatment is worth two times that in the low-incentive treatment.

One possible reason why subjects do not spend more time could be that they do not know the stakes. In our experiment, as in real-life insurance decisions, without calculating the variance of the

[1] Loewenstein et al. (2013) shows that consumers choose dominated plans, which are worse than another available plan regardless of preferences.

[2] Total health costs are the sums of the premiums and the costs of using the medical services, called "out-of-pocket" costs.

[3] We use the time to proxy for effort because careful decisions take time. However, time is not a perfect measure. We discuss these shortcomings further in the design (section 3) and the results (section 4).

total health costs, subjects may fail to realize how much their mistakes matter. If they do not know that they are in a high-stakes environment, they may not put in the effort. To test this hypothesis, within each incentive level, we implement another treatment, disclosure, in which subjects are told the stakes before they choose insurance plans. We find that disclosure does not change the results: knowing the underlying stakes does not impact the time spent. We, thus, conclude that subjects do not spend more time deciding because they perceive the returns of effort to be small, or at least smaller than the increase in incentives.

Although effort does not change with higher incentives, surprisingly, subjects do better, improving the number of correct answers by 0.28 from the base of 1.4 correct answers (out of 5). This hints at an under-explored alternative channel through which incentives affect choices, reducing difficulty. When incentives increase, the difference in the total health costs is larger, and hence, it may be easier to tell a good plan from a bad one.[4] While the results so far point towards this channel, we cannot disentangle difficulty from high incentives in this design. We conducted a follow-up experiment to do so.[5] This introduced a "low-easy" treatment, which scales the high-incentive (or "easy") questions using a new exchange rate between the experimental points and the dollar amounts subjects receive. Doing so achieves the low-incentive payment but keeps the "easy" structure the same. We find that subjects in the "low-easy" treatment do better without spending more time than those in the original low-incentive treatment. As such, we confirm that high-incentive problems are indeed easier.

In summary, we identify two potential channels by which incentives affect performance: increasing effort and reducing difficulty. We find no evidence for the former channel but evidence for the latter, which is surprising. Both are general effects of incentives, but the latter channel has received less attention. Both of these channels could interact, separately or together, with education.

Next, we study the impact of financial education on decision processes. We do so by randomly making education available to some subjects. These subjects have access to worked-out examples illustrating the steps to choose a good plan through a series of buttons at the end of each question. By measuring the time lapse between button clicks, we can measure the extent to which subjects use the materials and whether that improves choices. Moreover, we randomize education with the incentive treatment to study the interaction between education and incentives.

A natural explanation for ineffective financial education is that people do not use the materials. We do not find this to be the case in our experiment. 32% of subjects with available education use the materials. Furthermore, because the overall time on the task increases, the use of the educational materials does not fully substitute time spent on methods that may have been used in the absence of education. We also find that providing education has a positive effect on performance, improving the number of correct answers by 0.22.

[4] For example, when a good plan costs \$800 and a bad plan costs \$810, subjects need only make a calculation error of \$10 to choose wrongly. However, if the plan costs \$900 instead, the calculation error has to \$100, which is less likely, for the subjects to choose the bad plan.

[5] A possible effect of high incentives could be effort intensity. Subjects may work harder in the same amount of time. The follow-up experiment removes this channel. We also discuss this channel further in the results in section 4.

We now turn to the interaction between education and incentives. One may expect that a high incentive encourages more educational effort, but we find that subjects in the low-incentive treatment spend as much time in the educational materials as those in the high-incentive treatment. Given the identified channels behind incentives, we know that the low-incentive questions are hard; so, subjects are willing to put in the educational effort for hard problems.[6] However, it is less clear what causes the educational effort in the "high-easy" treatment. The follow-up experiment pins down the reason to incentives, and not ease: subjects in the "low-easy" treatment do not put in the time to read the materials. As a result, we conclude that subjects use educational materials either when problems are hard or when mistakes are highly costly.

As effort increases equally for both incentive levels, i.e., for both "high-easy" and "low-hard" problems, the interaction between education and incentives sheds light on whether the benefits of educational effort differ across difficulty levels. We find that providing education improves performance for easy questions but not for hard ones. In other words, the return of education is zero for hard problems, but it is potentially positive for easy problems.

To further understand subjects' choices, we also look for heterogeneous effects and the type of mistakes subjects make. Using a separate calculation task to classify subjects, we find that those who do well in the calculation task also do well in the insurance task. They are also better at improving their choices when the questions are easy. On the type of mistakes, we check if a dollar increase in premiums has the same weight as a dollar increase in "out-of-pocket" costs. Our evidence suggests that this is not the case: subjects place more weight on premiums. Providing education also increases subjects' sensitivity to premiums more, underscoring that education is more effective for easier problems.

Overall, there are three main findings of this project. First, high incentives have a surprising alternative channel in making the problems simpler. Second, subjects perceive education to be beneficial in either hard problems or high-stakes environment. Third, the actual return of education is positive for easy problems but zero for hard problems. The combination of the last two findings suggests that in financial choices, people could be encouraged to use education when stakes are high, but policy-makers should aim to reduce the difficulty of the problems so that the educational effort translates to better choices.

The above results were derived from two experiments. The main experiment recruited 2009 subjects, and the follow-up experiment recruited another 603 subjects. We paid subjects a participation fee and a bonus based on their choice. In the main experiment, we randomized subjects into eight treatments from a $2 \times 2 \times 2$ design, incentive $\times$ disclosure $\times$ available education. In the follow-up experiment, there were three treatments: "low-hard-no education", "low-easy-no education", and "low-easy-education". At the end of the experiment, subjects saw their bonus, and we transferred the payment to their MTurk account.

This project contributes to several strands of the literature. Other experimental papers have

[6] Low incentives could not have motivated subjects. So the educational effort must have been due to task difficulty.

used hypothetical insurance choices to study financial decisions (Johnson et al. (2013); Bhargava et al. (2015)). However, neither of them manipulate incentives by changing the features of the plans or study the effects of incentives on effort and the interaction between incentives and education. By manipulating incentives via changing premiums, our paper shows that high incentives do not affect effort, but they still matter by reducing difficulty. We also find an interaction between incentives and education on improving choices.

In the broader consumer finance literature, there is much interest in both disclosure and education. Existing papers disclose incentives by translating financial concepts (for instance, interest rate) into dollar amounts (Bertrand and Morse (2011); Goda et al. (2014)). We disclose incentives by showing the dollar difference between the best option and a randomly chosen option. On financial education, the literature, despite its size, has not said much on the factors contributing to education effectiveness. Our experiment shows that varying the structure of the products, premiums in our case, can complement education.

For the rest of this paper, we proceed with a simple framework in section 2, which highlights the mechanisms of incentives and education. We then show how we vary these three elements in our experiment in section 3. Section 4 presents our results from the experiment. Section 5 details the follow-up experiment and its results. Finally, we relate this paper to related literature in section 6 and conclude in section 7.

2 Framework

This section presents a framework capturing the key elements in the decision environment: incentive, disclosure, and education availability, which we map to the health insurance task. Then, using the framework, we show that measuring effort identifies the channels through which incentives and education have an effect.

2.1 Setting

Consider a decision-maker (DM) i who chooses from a list of insurance plans. This choice can be of high-stakes or low-stakes: the difference between the best plan and a random choice can be large or small. The stakes are denoted as an unknown $s \in \{H, L\}$ with a known prior $\mathbf{P}(s = H) = \mu$.[7] These stakes can be disclosed or undisclosed. We denote disclosure as $d \in \{U, D\}$. Educational materials may or may not be available, denoted as $l \in \{0, 1\}$.

The above three elements, s, d, and l, feature in the DM's timeline to choose a plan as follows:

1. The DM forms her belief of the stakes, $\hat{s}_i(s, d) = \mathbf{P}_i(s = H)$
2. The DM decides how much effort, e_i, and how much educational effort, e_{il} to spend

[7] The experiment uses a simple uniform prior: 50% high-stakes and 50% low-stakes.

3. The DM receives the result of her choice, $f_i(e_i; s)$

We explain each of the above steps in turn. Before attempting the choice, the DM forms her belief of s, $\hat{s}_i(s,d)$. If she is in the disclosed treatment, $d = D$, we display s transparently. As a result, the DM's belief is degenerate and correct: $\hat{s}_i(H,D) = 1$ and $\hat{s}_i(L,D) = 0$. If the DM is in the undisclosed treatment, $d = U$, we do not give her any other information about s except the prior μ. She may examine and compare the plans to move her belief (correctly or incorrectly) towards either H or L. She may decide not to do so and maintain the belief at μ. In any of the cases, $\hat{s}_i(\cdot)$ is the DM's belief before she makes any decisions.

The DM then decides how much effort $e_i \geq 0$ to choose a plan. If education is available, $l = 1$, then e_i may contain $e_{il} \geq 0$, the effort put into studying the educational materials. Formally, we decompose e_i as $e_i = e_{il} + e_{in}$, where e_{il} is the educational efforts and e_{in} includes all other types of effort. Note that when education is unavailable, $l = 0$, then $e_{il} = 0$.

The DM's effort translates to the number of correct answers, $f_i(e_i; s) = f_i(e_{il}, e_{in}; s)$ where $f_i(e_{il}, e_{in}; \cdot)$ is concave in each of the component of effort. We allow s to affect the number of correct answers (conditional on the same level of effort) because, under high stakes, the plan costs are further apart, so it may be simpler to tell a good plan from a bad one. For example, if the DM estimates the costs of the plans, under high stakes, she needs to make a large estimation error to confuse the relative quality of the plans. Meanwhile, under low stakes, she need only make a small error to choose the wrong plan. s can, thus, affect f_i directly.

The DM's optimization problem is:

$$\begin{aligned} &\max_{e_{il},e_{in}} \mathbf{E}_{\hat{s}_i(\cdot)}(s) f_i(e_{il}, e_{in}; s) - (e_{il} + e_{in}) \\ \text{subject to} \quad & e_{il} = 0 \text{ if } l = 0 \end{aligned}$$

The DM perceives the return to her effort to be her task performance $f_i(\cdot)$ multiplied by the expected reward of doing well, which is the expected stakes under her belief $\hat{s}_i(\cdot)$. Since $f_i(e_{il}, e_{in}, \cdot)$ is concave, we can assume that the cost of effort is linear, without loss of generality. Let $e_i^*(s,d,l) = (e_{il}^*(s,d,l), e_{in}^*(s,d,l))$ be the DM's choice, which satisfies

$$\mathbf{E}_{\hat{s}_i(\cdot)}(s) \frac{\partial f_i(e_i^*(\cdot); s)}{\partial e_i^*} = 1 \tag{1}$$

Because we assume that $f(\cdot)$ is concave, then an increase in $\hat{s}_i(\cdot)$, i.e., a greater belief that the choice is high-stakes, leads to an increase in $e_i^*(\cdot)$. Note that the choices of $e_{il}^*(\cdot)$ and $e_{in}^*(\cdot)$ satisfy the same condition. With this framework, we can study the effects of incentives and education.

2.2 Effort Identifies the Mechanisms of Incentives and Education

We look at the roles of incentives and education in turn. Using our setting, we can decompose the effect of incentives on performance, $f_i(\cdot)$ into two components as follows:

$$\begin{aligned}\Delta f_i(L \to H, \cdot) = & \quad f_i(e_i^*(H,\cdot),H) - f_i(e_i^*(L,\cdot),L) \\ = & \quad \underbrace{(f_i(e_i^*(H,\cdot),H) - f_i(e_i^*(L,\cdot),H))}_{\text{Effort}} + \underbrace{(f_i(e_i^*(L,\cdot),H) - f_i(e_i^*(L,\cdot),L))}_{\text{Difficulty}} \end{aligned} \tag{2}$$

Equation (2) show that without measuring incentives, we have an identification problem when $\Delta f_i(\cdot)$ is positive. If incentives affect choices, there can be two explanations: the DM increasing effort or the problem becoming simpler. It is natural to lean toward the former explanation. But equation (2) clarifies that this may not be the case; indeed, as we shall see in our results in section 4, difficulty explains a positive $\Delta f_i(L \to H, \cdot)$.

When the DM's effort does not increase with s, from the optimization condition in equation 1, we can attribute the lack of effort response to two cases. In the first case, $\mathbf{E}_{\hat{s}_i(\cdot)}(s) \neq s$; although s increases, $\hat{s}_i(s,d)$ may not move towards H accordingly or at all. To evaluate this reason, we can compare the efforts when stakes are disclosed, i.e., comparing $e_{il}^*(L,D,\cdot)$ with $e_{il}^*(H,D,\cdot)$. If $e_{il}^*(H,D,\cdot) > e_{il}^*(L,D,\cdot)$, then the DM does not increase effort because she does not realize that she is in a high-stakes environment even when $s = H$. Otherwise, if $e_{il}^*(H,D,\cdot) = e_{il}^*(L,D,\cdot)$, then the DM does not increase effort because of the second case: $\frac{\partial f_i(\cdot)}{\partial e_i}$, or the returns of effort, is low.

We now turn to the effect of education. If the DM does not put in the educational effort, it is hard to see how education may have an effect. So, we consider the effects of education on choices only when $e_{il}^*(\cdot) > 0$. When the DM uses the materials, three scenarios can happen to the overall effort, each of which corresponds to a different relation between e_{il} and e_{in}. First, if the cross derivative $\frac{\partial^2 f_i(\cdot)}{\partial e_{il}\partial e_{in}} < -1$, the educational effort reduces the usefulness of other methods more than proportionately, i.e., e_{il} more than substitutes e_{in}. As a result, $e_i^*(\cdot)$ decreases, i.e., education saves effort. The return to education then needs to take into account this reduction in effort besides any change in $\Delta f_i(0 \to 1, \cdot)$. Second, if $\frac{\partial^2 f_i(\cdot)}{\partial e_{il}\partial e_{in}} = -1$, then e_{il} exactly substitutes e_{in}; so, any changes in $f_i(\cdot)$ evaluate the relative effectiveness of e_{il} and e_{in}. In the last case, $\frac{\partial^2 f_i(\cdot)}{\partial e_{il}\partial e_{in}} > -1$, e_{il}^* does not fully substitute e_{in}^* and $e_i^*(\cdot)$ increases.

Finally, given that we study incentives and education, we can investigate whether their interaction changes behaviors. Intuitively, we expect high incentives encourages more educational effort, i.e., $\Delta e_{il}^*(0 \to 1, H, \cdot) > \Delta e_{il}^*(0 \to 1, L, \cdot)$. However, because s can affect how difficult the choice is, it is possible that the reverse is true: the DM believes that when the choice is more difficult, educational effort is more beneficial. In either case, if the interaction affects effort, we can ask whether the higher effort is associated with more effective education.

In summary, measuring efforts allow us to identify the channels through which incentives (effort versus difficulty) and education (educational materials substituting other methods or not) affect choices. Specifically, we use this framework to guide our analysis through the following questions:

1. Does effort increase with incentives, $e_i^*(H,\cdot) > e_i^*(L,\cdot)$?
   If it does, change in performance is associated with the DM working harder. If it does not, change in performance is associated with the problem becoming simpler.

2. If effort does not increase with incentives, does it increase with *disclosed* incentives, $e_i^*(H,D,\cdot) > e_i^*(L,D,\cdot)$?
   If it does, then the lack of effort response is due to wrong perceived reward to doing well. If it does not, then the lack of effort response is due to low returns of effort.

3. Does performance change with incentives, $\Delta f_i(L \to H,\cdot)$?

4. Does the DM use educational materials, $e_{il}^* > 0$? If the DM uses the materials, does providing education change the overall effort, e_i^*?
   The combination of changes in $e_{il}^*(\cdot)$ and $e_i^*(\cdot)$ reveal whether the educational materials substitute or complement other methods of choosing an insurance plan.mf

5. Does providing education improve performance, $\Delta f_i(0 \to 1)$?

6. Does the interaction between education and incentives affect effort and performance, $\Delta e_{il}^*(0 \to 1,H,\cdot) > \Delta e_{il}^*(0 \to 1,L,\cdot)$ and $\Delta f_i(0 \to 1,H,\cdot) > \Delta f_i(0 \to 1,L,\cdot)$?

3 Experimental Design

This section details our experimental design. We first outline the experimental setting: our subjects, their tasks, and their compensation. We then describe the treatments in the main experiment and the follow-up experiment. Finally, we describe the data we collect for performance and effort.

3.1 Decision Environment: Subjects, Tasks and Payment

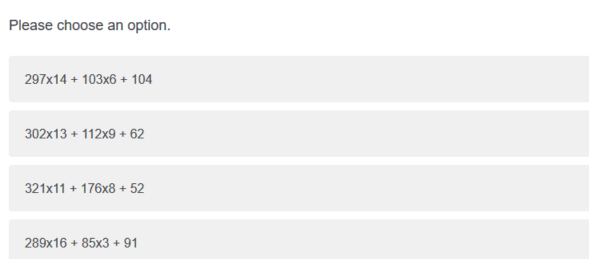
We recruited participants from Amazon Mechanical Turk (MTurk), which is a platform used by many social science experiments seeking a more representative population than university students. We restricted the subject pool to US workers because we would like the subjects to be familiar with the US health plan structure, which we use to design our plans. We posted the experiment as a Human Intelligence Task (HIT). Those who accepted the HIT followed a link to the experiment designed in Qualtrics, an online survey platform. For compensation, we paid them a participation fee of $2 and a bonus based on their choices in the experiment. The bonus is designed to incentivize subjects to spend effort, as further illustrated below.

As MTurk is an online platform, there is a worry that bots, instead of human workers, participated in the experiment. To minimize this concern, we restricted the subject pool to those who have completed more than 1,000 tasks and with approval ratings of more than 95%. Besides, workers needed to pass a captcha before entering in our experimental page. We discuss further this concern in our results in section 4.

Subjects completed two tasks: a calculation task, and then, a health insurance task. Since we expect subjects to base their insurance choices on arithmetic estimations, we use the calculation task to understand the subjects' baseline motivation and skills. We also classify subjects based on subjects' performance in the calculation task to measure heterogeneous treatment effects. At the end of the experiment, we choose one question from each task randomly and convert subjects' choices to the bonus.

We describe each task in turn. In the calculation task, there are four questions, an example of which is in figure 3.1. Each question contains four options, each option a sum. Subjects choose one sum, which earns points amounting to 5,000 minus the chosen sum. For example, for the question in figure 3.1, the first sum is 4,880. If a subject picks this sum, her points are $5{,}000 - 4{,}880 = 120$. The bonus payment is then 1 cent for each point. So, picking the first sum earns \$1.20 if the question in figure 3.1 is chosen for payment. [8] In this way, a subject earns the most points, and hence, the most money if she picks the smallest sum.

Figure 3.1: Calculation Task



After the calculation task, subjects complete the main task, a health insurance task. This task has five questions, an example of which is in figure 3.2. 3.2a zooms into the structure of the question. First, there is a hypothetical deterministic health-care scenario. Second, there are four plans whose structure mimics a US health insurance plan with a deductible, a co-payment/co-insurance, and a maximum out-of-pocket cost.[9] Subjects choose a plan for the scenario. Because the health care scenario is deterministic, the costs of all plans are deterministic. The lowest-cost plan is the objectively correct answer. Subjects' points in this task equal 10,000 minus the total cost of the chosen plan. So, they are incentivized to choose the lowest-cost plan, which matches real-life decisions. This payment scheme explains the scheme for the calculation task: we would

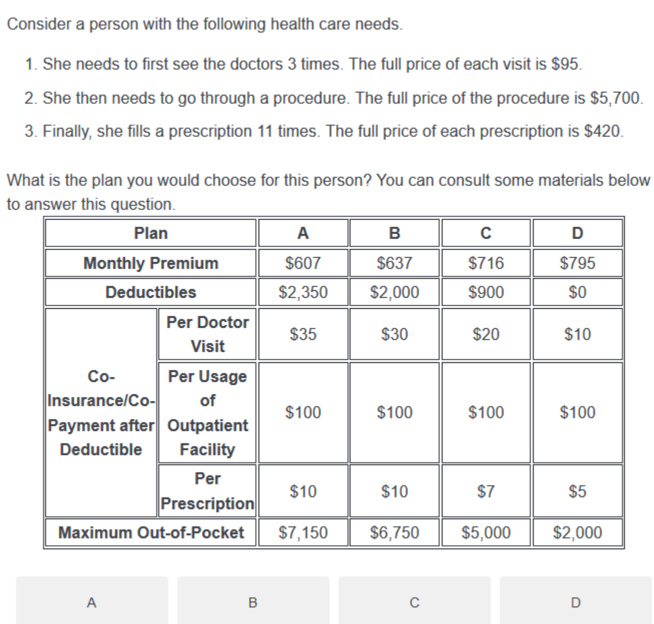
[8] There are step-by-step instructions and comprehension questions of how the bonus payment works before each task. Subjects cannot proceed to the task without passing the comprehension questions.

[9] Appendix A.3 provides more information on these terms.

like to maintain consistency in how we pay subjects to minimize confusion.

Figure 3.2: Health Insurance Task

(a) Question



Consider a person with the following health care needs.

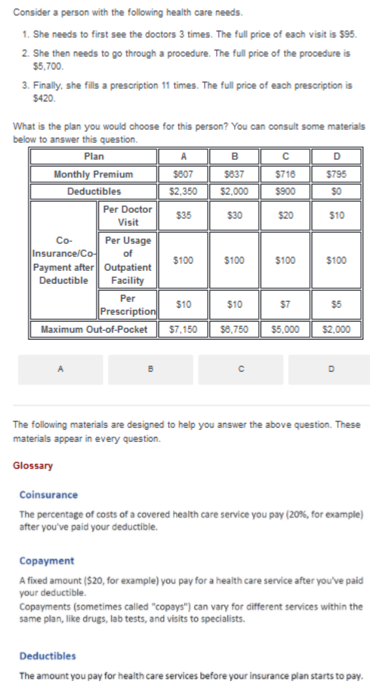
1. She needs to first see the doctors 3 times. The full price of each visit is $95.
2. She then needs to go through a procedure. The full price of the procedure is $5,700.
3. Finally, she fills a prescription 11 times. The full price of each prescription is $420.

What is the plan you would choose for this person? You can consult some materials below to answer this question.

Plan		A	B	C	D
Monthly Premium		$607	$637	$716	$795
Deductibles		$2,350	$2,000	$900	$0
Co-Insurance/Co-Payment after Deductible	Per Doctor Visit	$35	$30	$20	$10
	Per Usage of Outpatient Facility	$100	$100	$100	$100
	Per Prescription	$10	$10	$7	$5
Maximum Out-of-Pocket		$7,150	$6,750	$5,000	$2,000

A B C D

(b) Question and Material



Consider a person with the following health care needs.

1. She needs to first see the doctors 3 times. The full price of each visit is $95.
2. She then needs to go through a procedure. The full price of the procedure is $5,700.
3. Finally, she fills a prescription 11 times. The full price of each prescription is $420.

What is the plan you would choose for this person? You can consult some materials below to answer this question.

Plan		A	B	C	D
Monthly Premium		$607	$637	$716	$795
Deductibles		$2,350	$2,000	$900	$0
Co-Insurance/Co-Payment after Deductible	Per Doctor Visit	$35	$30	$20	$10
	Per Usage of Outpatient Facility	$100	$100	$100	$100
	Per Prescription	$10	$10	$7	$5
Maximum Out-of-Pocket		$7,150	$6,750	$5,000	$2,000

A B C D

The following materials are designed to help you answer the above question. These materials appear in every question.

Glossary

Coinsurance

The percentage of costs of a covered health care service you pay (20%, for example) after you've paid your deductible.

Copayment

A fixed amount ($20, for example) you pay for a health care service after you've paid your deductible.
Copayments (sometimes called "copays") can vary for different services within the same plan, like drugs, lab tests, and visits to specialists.

Deductibles

The amount you pay for health care services before your insurance plan starts to pay.

Each question in the insurance task has accompanying materials to help subjects choose a plan. The materials always include glossary definitions which are the standard definitions available with any real-life plan. Figure 3.2b shows a complete screenshot of a question followed by the materials.

At the end of the experiment, we ask subjects debriefing questions. For example, we ask subjects what they think is the stakes underlying the questions. Note that this debriefing happens *before* subjects know their final payout, so their answers are not affected by potential feedback.[10]

3.2 Treatments in Main Experiment

The experimental treatments apply to only the health insurance task. To minimize confusion and spillovers across treatments, we use a between-subject design.[11] Subjects are randomized into $2 \times 2 \times 2$ (high versus low incentive $\times$ no disclosure versus disclosure $\times$ no education versus education) treatment cells summarized in table 3.1.

The health insurance task has three components corresponding to each treatment.

1. Incentive display: corresponds to disclosure treatment

[10] The complete experiment, in either document format or Qualtrics format, is available upon request.

[11] It is confusing to have low and high-incentive questions alternate while it is hard to identify a clean treatment effect for education availability in a within-subject design.

Table 3.1: Between-subject Treatment

		No Education	Education
Low	Undisclosed	LU0	LU1
	Disclosed	LD0	LD1
High	Undisclosed	HU0	HU1
	Disclosed	HD0	HD1

2. Question: corresponds to incentive treatment

3. Accompanied materials: correspond to education treatment

As the question is the main component of the task, we explain this component first and then show how the other two components support answering the questions. We vary the incentives in the questions by changing the premiums of the plans while keeping the scenario and all other features of the plans the same. Figure 3.3, highlights the monthly premium row, the only difference across incentive levels. There are two reasons to focus on altering premiums instead of other features to increase stakes. First, changing the premium maintains the structure of the questions: the "out-of-pocket costs" of the plans are the same, both in the amount and the calculation method.[12] In other words, the more complicated part of finding the total health costs, which requires subjects to compare the health care needs with deductibles and co-payments or co-insurance, is the same across the incentives. The treatments differ in the simpler part: which number needs to be multiplied by 12 to find the yearly price of health insurance. The second reason to use premiums is that when we survey the plans in the market, plans across companies are often the same in their features except for the premiums.[13] As a result, we keep the same plan structure in the market and vary only the premiums.

Note that our design of the incentive treatment differs from the standard method of manipulating incentives. In most decision experiments, incentives change because the exchange rate between experimental points and bonus changes (Johnson et al. (2013),Dewan and Neligh (2017)). For instance, 1 point can be converted to either 1 cent or 2 cents, and the 1-cent treatment is of low incentive. We use this exchange-rate method in our follow-up experiment. However, in the main experiment, we vary the incentives by changing the points of the plans and keeping the exchange rate constant because this is how plans are presented in the real world. In real-life choices, "exchange rate" is always the same as there is only one currency, but how much plans cost in that currency can change. When that happens, subjects *can* figure out the incentives, but they *may not* do so. As a result, the effects of incentives may be diminished because subjects do not know the stakes. By mimicking incentives in real-world decisions, we can then ask if *undisclosed* stakes affect choices less than *disclosed* stakes.

[12] The "out-of-pocket costs" refer to what a patient pays to use medical services outside of the premiums.

[13] There are also other non-monetary differences across companies, for example, in-network services, and perhaps quality. These dimensions are outside the scope of the experiment.

Figure 3.3: Incentive Treatment

(a) Low Incentive

Consider a person with the following health care needs.

1. She needs to first see the doctors 3 times. The full price of each visit is $95.
2. She then needs to go through a procedure. The full price of the procedure is $5,70
3. Finally, she fills a prescription 11 times. The full price of each prescription is $420.

What is the plan you would choose for this person? You can consult some materials bel to answer this question.

Plan		A	B	C	D
Monthly Premium		$607	$637	$716	$795
Deductibles		$2,350	$2,000	$900	$0
Co-Insurance/Co-Payment after Deductible	Per Doctor Visit	$35	$30	$20	$10
	Per Usage of Outpatient Facility	$100	$100	$100	$100
	Per Prescription	$10	$10	$7	$5
Maximum Out-of-Pocket		$7,150	$6,750	$5,000	$2,000

(b) High Incentive

Consider a person with the following health care needs.

1. She needs to first see the doctors 3 times. The full price of each visit is $95.
2. She then needs to go through a procedure. The full price of the procedure is $5,700.
3. Finally, she fills a prescription 11 times. The full price of each prescription is $420.

What is the plan you would choose for this person? You can consult some materials below to answer this question.

Plan		A	B	C	D
Monthly Premium		$615	$645	$686	$811
Deductibles		$2,350	$2,000	$900	$0
Co-Insurance/Co-Payment after Deductible	Per Doctor Visit	$35	$30	$20	$10
	Per Usage of Outpatient Facility	$100	$100	$100	$100
	Per Prescription	$10	$10	$7	$5
Maximum Out-of-Pocket		$7,150	$6,750	$5,000	$2,000

We alter the premiums such that in either incentive treatments, choosing a plan randomly will earn subjects $2.25 on average. Under the low incentive, always choosing the best plan gives subjects on average $3.50 while under the high incentive, the best plan on average yields $7. The payment from the best plan in the high-incentive treatment doubles that of the low-incentive treatment. Given the average wage on MTurk, this difference is a considerable amount, warranting labeling the $7 treatment as "high incentive".[14]

To vary disclosure, before the questions, we randomize subjects to see different screens informing them of the incentive levels. Under no disclosure, subjects see the prior distribution (50% chance they are in either treatment). Under disclosure, subjects see the specific incentive to which they have been assigned. Figure 3.4 shows the difference across disclosure treatments conditional on the incentive level being low.

Figure 3.4: Disclosure Treatment, Conditional on Low Incentive

(a) Undisclosed Treatment

If you **choose a plan randomly**, you will receive **on average $2.25**.

In this section, there is a **50% chance** that you are in group X and a **50% chance** that you are in group Y.

If you are in **group X**, the **lowest-cost plan** gives you **on average $3.5**.

If you are in **group Y**, the **lowest-cost plan** gives you **on average $7**.

(b) Disclosed Treatment

If you **choose a plan randomly**, you will receive **on average $2.25**.

If you always choose the **lowest-cost plan**, you will receive **on average $3.5**.

[14]Hara et al. (2018), which looks at the payment and time worked on HITs from more than 2,000 workers, finds that, on average, workers earn $2/hour. This low wage is due to a small number of "bad" requesters who release a large number of lowly paid tasks. The average requesters paid $11/hour. Using these wage rates, $3.50 is worth between 19 to 105 minutes of a worker's time.

To vary available education, we randomize subjects to receive additional worked-out examples in the materials at the end of each questions. Although all subjects have access to glossary definitions, we note that this information does not show subjects the process of figuring out the right choice. In contrast, the examples show all the necessary steps to solve problems similar to those subjects have to answer. In other words, we devise deterministic health scenarios similar to those in the questions, and guide subjects on the cost calculation for two sample health plans. These extra materials are accessible via a series of buttons, so subjects can choose to use the examples or not.

3.3 Treatments in Follow-Up Experiment

In anticipating that our main experiment points to the conclusion that high-incentive questions are easier, we describe how the follow-up experiment is designed to clarify the effects of incentive. For a clearer explanation, we modify the names of the treatments in the main experiment. Because incentives may have two parallel channels, encouraging effort and reducing difficulty, the full name of "high incentive" should be "high-easy", and that of "low incentive" should be "low-hard".

To disentangle the two channels, we introduce the "low-easy" treatments by scaling the payments of the "high-easy" treatment. Specifically, the "low-easy" treatment has the same questions as the "high-easy" treatment. With the same scenarios and the same plans, the points of the plans, which are 10,000 minus the total health costs, are the same. To get a "low" payment, the follow-up exchange rate is 1 point to 0.25 cents instead of 1 point to 1 cent in the main design. While this new exchange rate gives the desired "low" difference between a randomly chosen plan and the best plan, it makes the overall payment too small. So, we also pay subjects a completion fee of $1.75, a payment they receive at the end of the insurance task regardless of their choices. In other words, if we apply only the new exchange rate, choosing randomly in the "easy" questions earns subjects $0.5 and always choosing the best plan gives $1.75. Adding the completion fee of $1.75 results in $2.25 for choosing randomly and $3.5 for always choosing the best plan, matching the low incentive in the main design.

As "low-easy" treatment pays the same as the "low-hard" treatment, comparing subjects' behaviors across these two treatments reveal the effects of easiness. To do so with or without education, we include both the "low-easy-no education" and "low-easy-education" treatments in the follow-up experiment. Besides, we include the "low-hard-no education" (LD0) as the "linking treatment", i.e., a treatment that appears in both the follow-up and the main experiments, to serve as a basis to compare results *across* experiments. If subjects from the linking group are similar across the two experiments, we can compare the follow-up treatments and the main treatments to find the relative importance of difficulty.

Table 3.2 summarizes the three treatments in the follow-up experiment, all of which have their counterparts in the main experiment. The linking group is the same as its counterpart while the other two is derived from the counterparts via the new exchange rate. Note that all follow-up treatments are disclosed and gives a "low" payment. This means that all subjects see the screen in figure 3.4b stating that random choice yields \$2.25 and the lowest-cost plan yields \$3.50 before the insurance questions.

Table 3.2: Between-Subject Treatment: Main vs. Follow-up

Follow-Up	Main
Low-Hard-NoEdu (LD0)	Low-Hard-NoEdu (LD0)
Low-Easy-NoEdu*	High-Easy-NoEdu (HD0)
Low-Easy-Edu*	High-Easy-Edu (HD1)

*: new exchange rate

In the follow-up experiment, subjects face the same environment as in the main experiment: answering questions on Qualtrics accessed via Amazon MTurk. They go through the same calculation task. When they reach the insurance task, one-third of the subjects (the linking group) see the 1-to-1 exchange rate while two-thirds see the 1-to-0.25 exchange rate. Among the two-thirds, half receive extra educational materials.

3.4 Measurement: Choice and Effort

We measure choice by the number of questions subjects choose the best plan. To proxy for overall effort, we use the amount of time subjects spent choosing insurance plans because careful decisions take time. To proxy for educational effort, we use the time lapse between button clicks in the educational materials. We ignore the first click to minimize capturing impulsive clicking. If subjects click on the second button, we consider the subjects to have used the materials. We use the time taken between the second click and the last click within the education section to measure education time. Since subjects can continue reading or processing the materials after the last click, we note that our measure is the lower bound of the actual time spent on education. Besides, we recognize that time misses effort intensity. We provide suggestive evidence that this is not a significant concern in the results (section 4), and address this shortcoming more explicitly in the follow-up experiment.

In summary, we collect the following data for performance and effort:

- $f_i(\cdot)$: the number of questions where subjects choose the best plan
- e_i^*: the total time subjects spend on the insurance task
- e_{il}^*: click data within the education section

We use the above data and apply the analysis from the framework in section 2 to understand the effects of incentives and education in subjects' behaviors.

4 Results: Main Experiment

All of the results in this section are from the main experiment described in section 3, although we interject with the results from the follow-up experiment in section 5 where appropriate. Before discussing the treatment effects, we give an overview of the subjects by providing descriptive statistics of their demographics and their performance in the calculation task, which classifies them into types. We proceed to discuss the effects of incentives, disclosure, and education availability. We wrap up with the heterogeneous effects using the types defined by the calculation tasks and the types of mistakes subjects make.

4.1 Description of Subjects

The main experiment collects 2,009 complete responses, which are distributed approximately evenly across treatments. Each treatment has between 249 to 253 responses.[15] On average, subjects completed the experiment in 1,413 seconds (24 minutes), earning \$0.7 from the calculation task and \$2.6 from the insurance task (on top of the \$2 participation fee).

There are no obvious concerns about selection bias. First, we randomly assign subjects into treatments. Second, we do not find attrition bias. Of the 123 incomplete responses, 90% abandon the experiment before the insurance task. The remaining 10% are present in all treatments. Third, we check for demographic balance across the treatments by checking for the treatment "effect" on the demographics, the result of which is in A.1. The only significant difference is that subjects randomized into the high-incentive treatment are less likely to have a health insurance plan, which is consistent with a 5% random chance of finding a significant difference.

There are three differences in demographics between the sample and the US population worth noting. [16] First, the sample is more educated, with 57% having a college degree or more, compared to 31% in the US Census. Second, they are younger: there are relatively few workers beyond the age of 40. Third, more of the sample, 18%, do not have health insurance compared to the 9% in the population. The last two differences agree with our prior of "gig workers" on an online platform. That the sample is relatively young possibly explains their higher education level. Although the differences with the US population are not essential to the study per se, it is useful to keep in mind that this sample is not representative of the consumers, and our results are local to this population.

4.2 Calculation Task

We give an overview of the subjects' performance on the task and then use their performance to classify them into two types. On average, subjects spend 4.1 minutes on the calculation questions, answering 1.94 questions correctly (out of 4). Although this performance is significantly better than

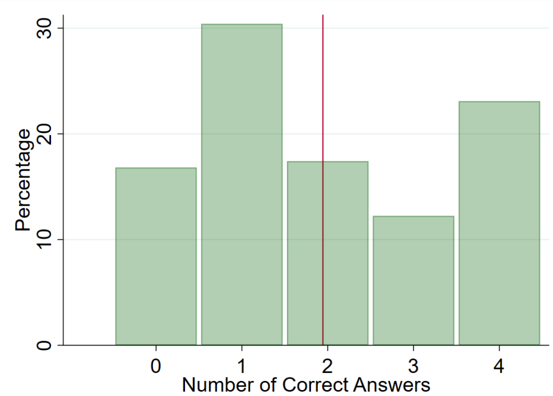
[15] The appendix presents information on the number of subjects per treatment.

[16] Appendix A.1 presents the details of the subjects' demographics as well as the number of subjects for each treatment arm.

randomization, at 1 correct answer, recall that the calculation task asks straightforward arithmetic questions, which subjects can complete however they wish, without a time limit.[17] So, even when subjects can answer the questions perfectly, the cost of doing so is non-trivial.

Figure 4.1 presents the distribution of subjects' performance, which shows a fair amount of heterogeneity. The distribution is spread out over all the possible number of correct answers, from 0 to 4 possible correct answers. The vertical line, at 1.94, indicates the average number of correct answers. As a simple classification of subjects, we label those who answer more than 1 question correctly, corresponding to doing better than randomly, as the "high type", θ^h, and the rest as θ^l. In our data, this classification happens to be a reasonably even split with 53% of subjects belonging to θ^h and 47% belonging to θ^l. We use this classification to understand heterogeneous effects in the main task.

Figure 4.1: Distribution of Subjects' Performance in Calculation Task



To provide a complete picture of the types, figure 4.2 presents the CDF of time θ^h and θ^l spend in different components of the experiment. Note that all analysis for time is done in logarithmic to correct for the heavy right tail in the time data.[18] Figures 4.2a and 4.2b show that θ^h generally spend more time to answer the questions in both tasks. Although that θ^l not spending time and not doing well may trigger the worry that they are bots spamming MTurk, we do not think this is the case. Figure 4.2c shows that there is no difference between the types in the time spent outside of the tasks, i.e., reading instructions. So, θ^l spend time to read instructions to understand the experiment, but when the questions appear, they do not spend time and answer them poorly.

4.3 Insurance Task

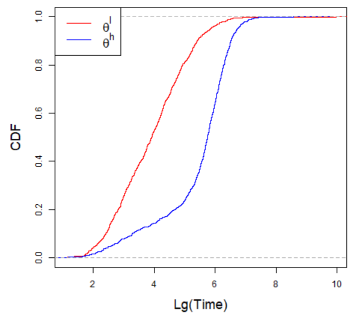
We give an overview of the subjects' performance before discussing the effects of incentives and education. On average, subjects spend 335 seconds (5.6 minutes) on the insurance task, answering

[17]Recall that there are four questions, and each question has four options. So randomization yields 1 correct answer.

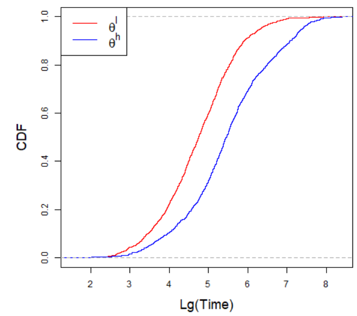
[18]Appendix A.2 presents the QQ-plots to contrast the distributions of the level of time and logarithmic of time. The logarithmic of time is closer to a normal distribution.

Figure 4.2: Overall Time Spent in Components of the Experiment by Subjects' Type

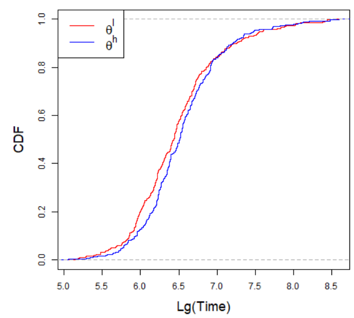
(a) CDF of lg(time) in Calculation Task



(b) CDF of lg(time) in Insurance Task



(c) CDF of lg(time) in Non-Task



1.6 questions correctly (out of five). There are variations in performance: subjects in LU0 treatment have the worst performance with 1.43 correct answers, and those in HD1 perform the best with 1.89. The performance in all treatments are significantly better than 1.25, the average accuracy under randomization.[19] 33% of subjects use the materials when they are available although they seem to spend a limited amount of time, only about half a minute, looking at the materials.[20] Appendix B.1 presents the summary statistics broken down by treatment arms.

We analyze the effects of incentives and education by measuring the treatment effects on effort and choices, as guided by the framework in section 2. The baseline specification is equation 3. In all of our results, we control for all treatments but sometimes suppress their coefficients for clearer exposition. On top of the baseline specification, we add interaction terms where appropriate.

$$y_i = \beta_0 + \beta_1 \mathbf{1}\{s = H\} + \beta_2 \mathbf{1}\{d = D\} + \beta_3 \mathbf{1}\{l = 1\} + \epsilon_i \tag{3}$$

[19] Recall that there are five questions in the insurance task, each with four options. So, the average number of correct answers under randomization is 1.25.

[20] The limited education time could be because our measure captures only the lower bound of time spent reading the materials. In a small test conducted before the main experiment to check that the educational materials can be understood, we recruit 30 workers to read the materials and answer ten comprehension questions based on the materials. In this test, on average, subjects spend 2.5 minutes on the materials and answer 7.3 questions correctly.

where

$$\begin{aligned}
y_i &\in \{f_i(\cdot), e_i^*, e_{il}^*\} \\
\mathbf{1}\{s = H\} &= \text{indicator for high incentives} \\
\mathbf{1}\{d = D\} &= \text{indicator for disclosure} \\
\mathbf{1}\{l = 1\} &= \text{indicator for available education}
\end{aligned}$$

4.3.1 Effects of High Incentives

As analyzed in the framework in section 2, we start with the impact of incentives on effort. Specifically, we focus on β_1, the coefficient of $\mathbf{1}\{s = H\}$ with y_i being the time spent on the task. Table 4.1 shows that high incentives have no effects on any of the three measures of effort, overall time, educational use, or educational time although the payment from the best plan in the high-incentive treatment doubles that in the low-incentive treatment. As it is surprising that subjects do not spend more time on high-incentive questions, we check whether our measure is reliable. Since subjects can finish the task however they wish without any time limit, one concern is that the time measure may be too noisy to detect any changes.[21] However, we find that the correlation between the logarithm of time and performance is 0.32, significant at 1%, so, the lack of effort response to incentives is unlikely due to noise alone.

Table 4.1: Effects of Incentives on Effort

	Lg(Time)	EduUse	Lg(EduTime)
High Incentive	0.0270	0.0167	-0.0993
	(0.0504)	(0.0296)	(0.137)
Constant	5.067***	0.296***	3.765***
	(0.0491)	(0.0250)	(0.130)
Observations	2009	1003	326

Standard errors in parentheses. * $p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.
The regressions control for the disclosure and education treatments.

Table 4.2 then examines whether the lack of response of time is caused by subjects not knowing how much their mistakes matter by looking at the coefficients on $\mathbf{1}\{s = H\}$, $\mathbf{1}\{d = D\}$ and their interaction. The null effect from the interaction indicates that this is not the reason. Appendix B.2 shows figures of the CDFs of overall time spent to highlight that the results are present in the entire distribution, and not just in the averages. From the analysis in section 2, we, thus, conclude that subjects perceive the return to efforts to be small.

[21] There is technically a time limit imposed on the MTurk interface: subjects have 3 hours between accepting the task and submitting the completion code. However, if they need more than 3 hours, they can enter their demographic information in the MTurk portal before 3 hours run out, and we match their demographic answers with the debriefing questions to pay them. All but five subjects completed the experiment within 3 hours.

Table 4.2: Effects of Incentives and Disclosure on Effort

	Lg(Time)	EduUse	Lg(EduTime)
High Incentive	0.0114	0.0456	-0.248
	(0.0707)	(0.0412)	(0.202)
Disclosure	-0.0529	0.0709*	-0.165
	(0.0712)	(0.0416)	(0.207)
High Incentive x Discl	0.0312	-0.0576	0.279
	(0.101)	(0.0592)	(0.274)
Constant	5.075***	0.281***	3.845***
	(0.0550)	(0.0285)	(0.158)
Observations	2009	1003	326

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
The regressions control for the education treatment.

We note here that table 4.2 shows that disclosure seems to affect educational usage when incentives are *low*. However, this occurs without an increase in educational time. Since educational use is defined as whether subjects access the materials, we attribute this effect to noises in subjects impulsively clicking the button.

Despite subjects not spending more time on high-incentive questions, they do better. Table 4.3 show that they answer more questions correctly, *regardless* of disclosure. The point estimates in the specifications with and without the interaction between disclosure and incentives are roughly the same. Subjects answer 0.27 questions more correctly in high-incentive questions, from an average of 1.36 correct answers in the low-incentive ones. Two hypotheses can explain this effect: effort intensity or easier choices. Within the same amount of time, subjects can be more alert and pay more attention. At the same time, since higher stakes make the total health costs further apart, they can make the problems simpler.

As we cannot fully disentangle the two hypotheses in the main experiment, we use the follow-up experiment to separate the channels. While the follow-up results in section 5 clarifies that high-incentive questions are easier, there is evidence in the main design which already hints that effort intensity is not the key reason because effort intensity can rationalize the results only under the *disclosed* treatments. In the *undisclosed* treatment, subjects increase effort intensity only if they know the stakes. However, their answers to our debriefing question suggest that they do not know the stakes. Table 4.4 tabulates the percentage of subjects choosing each available answer when they are asked what they think is the payment for the best plan. The majority of subjects in the disclosed treatments answer correctly, while the majority of subjects in the undisclosed treatments pick the prior. This table does not show the split across education availability, but the result, in appendix B.2, is the same. As a result, effort intensity may not justify the impact of incentives for undisclosed incentives.

Table 4.3: Effects of Incentive and Disclosure on Performance

	Number of Correct Ans	
High Incentive	0.273***	0.280***
	(0.0506)	(0.0714)
Disclosure	0.0350	0.0427
	(0.0506)	(0.0646)
High Incentive x Discl		-0.0154
		(0.101)
Constant	1.368***	1.364***
	(0.0487)	(0.0530)
Observations	2009	2009

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
The regressions control for the education treatment.

Table 4.4: Incentive Perception - % of Subjects by Their Answers

	Undisclosed		Disclosed	
	Low	High	Low	High
\$2.25	7.8	7.6	15.7	11.0
\$3.5	11.2	11.8	76.7	9.8
\$7	4.4	5.2	5.4	75.1
Prior: (50%: \$3.5, 50%: \$7)	76.7	75.5	2.2	4.2
Total	100	100	100	100

Q: If you always choose the lowest-cost plan, what do you think you earn? Correct Answer for L: \$3.50, for H: \$7

In summary, the framework in section 2 suggests that incentives work either by increasing effort or reducing difficulty. Our data finds no evidence for the former channel, but finds evidence for the latter. The follow-up experiment confirms this finding, which is surprising and unexplored in the literature. We now examine the impact of available education and then how the channels of incentives may interact with education.

4.3.2 Effects of Providing Education

Similar to the previous section, we first look at the effects of incentives on effort and then on choices. With 32% of subjects using the education materials, table 4.5, which focuses the coefficients on $\mathbf{1}\{l = 1\}$, $\mathbf{1}\{s = H\}$, and their interaction, shows that providing education materials increases overall time spent by more than 20%, suggesting that studying the materials does not fully substitute time spent on other methods which may have otherwise been used (for example, asking someone for help). The table also shows that the interaction between education and incentives has no effect on time spent:

subjects in low-incentive treatment spend as much time as those in high-incentive treatment. Since low incentives could not have motivated subjects in the low-incentive treatment, their educational effort must have been caused by task difficulty. Although we cannot disentangle the causes for educational effort for the high-incentive treatment, the follow-up experiment shows that incentives, not ease, encourage educational effort.

Table 4.5: Effects of Education and Incentives on Effort

	Lg(Time)	Lg(Time)
Education	0.216***	0.181**
	(0.0504)	(0.0712)
High Incentive	0.0270	-0.00792
	(0.0504)	(0.0700)
High Incentive x Edu		0.0699
		(0.101)
Constant	5.067***	5.085***
	(0.0491)	(0.0545)
Observations	2009	2009

Standard errors in parentheses. * $p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.
The regressions control for the disclosure treatment.

Since subjects increase effort equally for both incentive levels, i.e., for both "low-hard" and "high-easy" problems, the interaction between education and incentives on choices measures the education effectiveness on different difficulty levels. Table 4.6 shows that education is ineffective for hard problems despite subjects' educational effort, but education is effective for easy problems, increasing the number of correct answers by 0.21 questions. The pooled effect of education is positive at 0.146. Appendix B.3 shows the CDFs of performance and time spent to confirm that the results on the interaction are not confined to averages but are present in the entire distribution.

To provide a complete picture of the effects of available education, we show suggestive evidence on the selection into education. First, the correlation between the logarithm of calculation time and education time is 0.3, significant at 1%.[22] Second, if we assume that those not using education behave "as if" the materials are absent, comparing their time with time by all subjects under unavailable education further clarifies the selection. In other words, if we assume that the presence of education does not change the behavior of those not using education, the difference in time comes from those who would have used education if it were available. Figure 4.3 agrees with the first piece of evidence: those who already put in effort use education more. The difference in the mean log(time) is 0.154, significant at 1%. Education availability reinforces willingness to spend efforts.

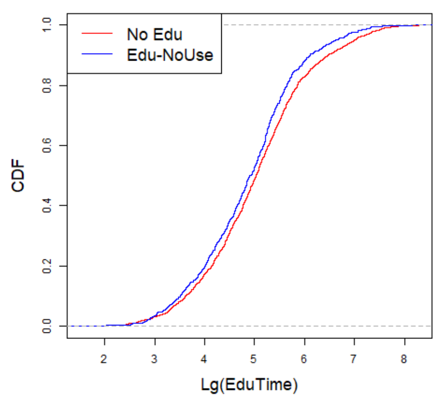
[22]The correlation between calculation performance and education time is 0.24, significant at 1%. This result is consistent with the evidence that those who do better in the calculation task spend more time.

Table 4.6: Effects of Education and Incentives on Performance

	Ans	Ans
Education	0.146**	0.0422
	(0.0506)	(0.0645)
High Incentive	0.273***	0.169**
	(0.0506)	(0.0685)
High Incentive x Edu		0.207**
		(0.101)
Constant	1.368***	1.419***
	(0.0487)	(0.0528)
Observations	2009	2009

Standard errors in parentheses. * $p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.
The regressions control for the disclosure treatment.

Figure 4.3: Overall Time - (Edu, No Use) versus No Edu



4.3.3 Heterogeneous Effects

After analyzing the treatment effects on the entire subject pool, we now look at the heterogeneous effects across the types categorized by the calculation task.

Similar to the difference in the calculation task, in the insurance task, θ^h's continue to spend more time and do better than θ^l's. We add an indicator for θ^h to the baseline specification 3 and show the differences between the types in 4.7. Moreover, figure 4.4 reiterates the selection into education: the high type is significantly more likely to use education.

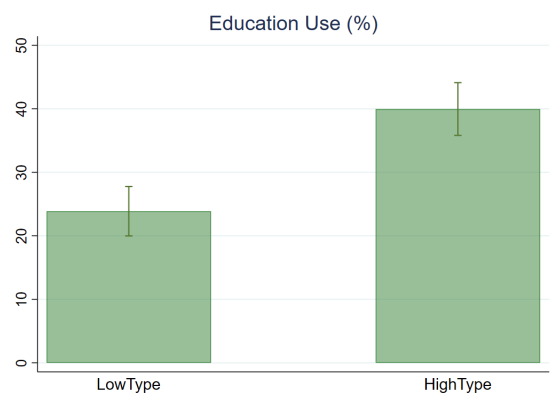
Table 4.8 shows the heterogeneous treatment effects by interacting the indicator for θ^h with that for high incentive and available education. Under high incentives, θ^h's do better, but the difference in time is insignificant. In other words, θ^h's effort response to a high incentive is not different from that of θ^l's, but the baseline longer time spent allows θ^h's to take better advantage of the easier

Table 4.7: Difference Between Types in Insurance Task

	Ans	Lg(Time)
θ^h	0.383***	0.776***
	(0.0492)	(0.0470)
High Incentive	0.273***	0.0282
	(0.0499)	(0.0474)
Education	0.139**	0.202***
	(0.0499)	(0.0474)
Constant	1.171***	4.668***
	(0.0535)	(0.0510)
Observations	2009	2009

Standard errors in parentheses. * $p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.
The regressions control for the disclosure treatment.

Figure 4.4: Difference Between Types in Education Use



questions.

When education is available, θ^h's neither spend more time nor make better choices. However, note that the coefficient on "education" for performance, which indicates the effect for θ^l, is now insignificant and is close to 0. Meanwhile, the point estimate for the interaction term, 0.125, is close to the overall effect of available education, 0.139, in table 4.7.[23] As a result, the statistically insignificant difference in performance is likely due to the lack of power, but we conclude that there is no difference in the educational benefit to the types.

Appendix B.4 shows the regression results with the full interaction between θ^h, incentive, and education. Here, we lack statistical power. The only significant difference that survives is the baseline difference between the types, the effect of incentive (alone) on performance, and the effect of available education (alone) on time spent.

[23] The sum of the coefficient on "education" and "education $\times\theta^h$" is significantly positive.

Table 4.8: Heterogeneous Treatment Effects

	Incentive		Education	
	Ans	Lg(Time)	Ans	Lg(Time)
θ^h	0.297***	0.710***	0.321***	0.699***
	(0.0633)	(0.0671)	(0.0673)	(0.0662)
High Incentive	0.183**	-0.0413	0.274***	0.0297
	(0.0615)	(0.0627)	(0.0499)	(0.0473)
Education	0.141**	0.204***	0.0728	0.121*
	(0.0498)	(0.0473)	(0.0616)	(0.0627)
High Incentive x θ^h	0.171*	0.132		
	(0.0985)	(0.0940)		
Edu x θ^h			0.125	0.154
			(0.0986)	(0.0939)
Constant	1.215***	4.702***	1.202***	4.706***
	(0.0551)	(0.0544)	(0.0553)	(0.0549)
Observations	2009	2009	2009	2009

Standard errors in parentheses. * $p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

The regressions control for the disclosure treatment.

4.3.4 Subjects' Mistakes

This section provides evidence of subjects' different sensitivities to cost components of the plans, premiums and "out-of-pocket" costs. We first explain how we use our data to tease out subjects' responses to the components and then present results of our estimates.

Our data set contains two essential details of subject-level choices: the entire menu of plans and the cost components of the plans. In other words, for each subject, we know the plans they choose and the plans they *do not* choose. Moreover, because we know the health scenarios attached to the plans, we know the premiums and the deterministic "out-of-pocket" costs of using medical services for each plan. As a result, we can find subjects' sensitivity to the cost components using a linear probability model specified in equation (4.)

$$
\begin{aligned}
y_{ijk} = \quad & \gamma_0 + \gamma_1 \text{Premium}_{jk} + \gamma_2 \text{OOP}_{jk} + \textstyle\sum_k \gamma_{3k} \mathbf{1}_k \\
& + \gamma_{1s} \text{Premium}_{jk} \times \mathbf{1}\{s = H\} + \gamma_{2s} \text{OOP}_{jk} \times \mathbf{1}\{s = H\} \\
& + \gamma_{1l} \text{Premium}_{jk} \times \mathbf{1}\{l = 1\} + \gamma_{2l} \text{OOP}_{jk} \times \mathbf{1}\{l = 1\} + \epsilon_{ijk}
\end{aligned} \tag{4}
$$

where

$$
\begin{aligned}
y_{ijk} &= \begin{cases} 1 & \text{if subject } i \text{ chooses plan } j \text{ in scenario } k \\ 0 & \text{otherwise} \end{cases} \\
\text{Premium}_{jk} &= \text{annual premium of plan } j \text{ in scenario } k \\
\text{OOP}_{jk} &= \text{out-of-pocket cost of plan } j \text{ in scenario } k \\
\mathbf{1}_k &= \text{indicator for scenario } k
\end{aligned}
$$

Table 4.9 shows the results with standard errors clustered at the subject level. The first column ignores the interactions with the treatment assignments while the rest check the treatment effects on the sensitivities. Overall, subjects respond to cost increase: an increase in either the premium or the out-of-pocket costs reduces the likelihood of choosing the plan, which reflects subjects' understanding of insurance. However, the response differs between cost components: subjects put more weight on premiums. The coefficients on premiums and "out-of-pocket" costs in the first column are different from each other at 1% statistical significance. Since premiums are much easier to perceive, it is reasonable that people are more responsive to this component. However, note that this is a "mistake" because in principle, subjects should be equally sensitive to the components. Our evidence is similar to observations documented in the literature (Johnson et al. (2013),Abaluck and Gruber (2011)).

Table 4.9: Sensitivity to Cost Components

	$\mathbf{1}$Choose Plan		
Premium	-0.00219***	-0.00262***	-0.00161***
	(0.000182)	(0.000381)	(0.000235)
OOP	-0.000319***	-0.000359***	-0.000272***
	(0.0000148)	(0.0000309)	(0.0000194)
Premium x High		0.000492	
		(0.000428)	
OOP x High		0.0000514	
		(0.0000345)	
Premium x Edu			-0.00116**
			(0.000356)
OOP x Edu			-0.0000932**
			(0.0000288)
Observations	40180	40180	40180

Standard errors in parentheses.

$^{*}\ p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Increasing the incentives via the premiums do not change either weight. Meanwhile, available education increases subjects' responses to both components although the increase is higher for premiums. The differential education effect on the components is surprising because a priori, given that premium is a relatively straightforward concept, we do not expect education to affect the weight attached to the premium. The evidence, thus, reinforces the earlier conclusion that available education is more effective on easier problems.

In summary, our main experiment suggests that incentive works because it makes the questions easier, which is a surprising channel. We also find although subjects are willing to put in the educational effort when the problems are hard but providing education works only under high-incentive or easy problems. On other results, we find that the difference between the types from the calculation task persists in the insurance task. Subjects also place more weights on premiums relative to "out-of-pocket" costs.

5 Results: Follow-up Experiment

The follow-up experiment is designed to clarify the effects of incentives and education detected in the main experiment. We start the follow-up results with a brief overview of the responses. We then provide evidence that subjects in the follow-up experiment are similar to those in the main experiment. This similarity allows us to pool the follow-up treatments with the main treatments for the analysis. Using the pooled data, we answer two questions from the main experiment. First, do high incentives make choices easier? Second, does ease or high incentive motivates educational effort?

There are 603 collected responses for the follow-up experiment, with 201 subjects in each treatment. On average, subjects completed the experiment in 1,291 seconds (22 minutes), earning \$0.68 from the calculation task and \$2.45 from the insurance task.

Across our measures of choice and effort, subjects from the follow-up experiment are not statistically different from those in the main experiment. Figure 5.1 shows that the follow-up subjects' performance in the calculation task is perceptibly indistinguishable from that in the main experiment. Here, the distribution is similarly spread out from 0 to 4 with 49% belonging to θ^l and 51% belong to θ^h.

Table 5.1, using an indicator for subjects from the follow-up experiment, provides further evidence that the two groups of subjects are similar. Overall, they are not different in their calculation performance. The linking treatment, LD0, is similar across two experiments in performance and time spent in the insurance task. Moreover, when education is available, 36% of subjects in the follow-up experiment use the materials, comparable to the 33% in the main experiment.

Given the similarity, we now pool the treatments from both experiments. Table 5.2, using an indicator variable for easy questions, show the effects of easiness with or without education. Without education, subjects in the "low-easy" treatment does better without spending more time than those in the "low-hard" treatment. Meanwhile, with education, subjects in the "low-easy"

Figure 5.1: Distribution of Performance in Calculation Task (Follow-Up)

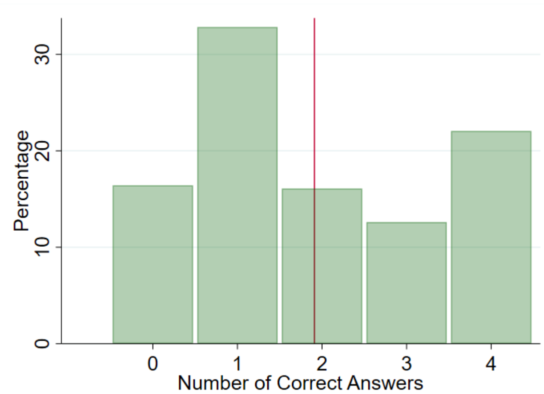


Table 5.1: Subjects in Main Experiments vs Subjects in Follow-Up Experiments

	All	LD0	
	Calc Ans	Ins Ans	Ins Lg(Time)
Follow-Up	-0.0333	-0.0447	-0.0401
	(0.0656)	(0.0914)	(0.101)
Constant	1.944***	1.443***	5.026***
	(0.0317)	(0.0627)	(0.0714)
Observations	2612	454	454

Standard errors in parentheses. $^{*}\ p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

treatment spend less time, and yet, do as well as those in the “low-hard” treatment. In either cases, the performance given an amount of time is higher for “low-easy” treatments, confirming that high-incentive questions are easier.

We now turn to the second question from the main experiment in table 5.3 which looks at the effects of providing education on “low-easy” subjects. When incentives are low and the problems are easy, subjects neither increase time nor performance. This suggests that when the problems are easy, removing the high incentive removes the motivation to put in the effort, which then explains the ineffectiveness of education. Table 5.4 provides further evidence of the effects of incentives when education is available and problems are easy. In this case, high incentives encourage educational effort, which, in turn, explains the effect of providing education on choices.

Recall that in the main experiment, we find that subjects are willing to put in the educational effort in the “low-hard” treatment. In contrast, we find in the follow-up experiment that they are not willing to do so for “low-easy” treatment. So, conditional on a low incentive, subjects perceive the returns of education to be higher for hard problems. However, the main experiment shows that this perception is wrong: subjects put in the same time for hard and easy problems but they improve performance only for easy problems. We leave the task of understanding why subjects have the wrong perception to later research.

Table 5.2: Effects of Easiness

	Low, Easy-Hard			
	No Edu		Edu	
	Ans	Lg(Time)	Ans	Lg(Time)
Easy	0.289**	-0.0964	0.0661	-0.247**
	(0.0958)	(0.0912)	(0.104)	(0.109)
Constant	1.423***	5.008***	1.516***	5.234***
	(0.0457)	(0.0509)	(0.0657)	(0.0716)
Observations	655	655	451	451

Standard errors in parentheses. * $p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 5.3: Effects of Education on “Low-Easy”

	Ans	Lg(Time)
Education	-0.129	0.0755
	(0.117)	(0.112)
Constant	1.711***	4.911***
	(0.0842)	(0.0758)
Observations	402	402

Standard errors in parentheses. * $p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

In summary, the follow-up experiment confirms that difficulty is critical in explaining the effect of incentives. When education is available, we find that ease and incentives complement each other.

6 Related Literature

This paper is related to a number of strands of literature. We discuss them from the specific literature on health insurance to the broadest literature on the effects of incentives in experiments.

The relatively recent attention of US policy-makers on health insurance has been matched with a number of studies using US data, which generally show that consumers are not choosing the financially optimal plan. Abaluck and Gruber (2011, 2016); Bhargava et al. (2015) use individual choices (under Medicare Part D for the first two papers, employer-sponsored health insurance for the third paper) to show that consumers can save a significant amount of money by making a different choice. Moreover, Abaluck and Gruber (2016), which follow up on Abaluck and Gruber (2011), reveals that consumers do not make better choices over time. Besides these observational studies on poor choices, Loewenstein et al. (2013) presents evidence that consumers have little understanding of health insurance.

There are a few experimental papers trying to understand health insurance choices. Kling et al.

Table 5.4: Incentives Complement Education for Easy Questions

	Easy,Edu,High-Low			
	Ans	Lg(Time)	EduUse	Lg(EduTime)
High Incentive	0.307**	0.309**	-0.0233	0.568**
	(0.116)	(0.109)	(0.0452)	(0.187)
Constant	1.582***	4.987***	0.363***	3.143***
	(0.0805)	(0.0824)	(0.0340)	(0.136)
Observations	454	454	454	159

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

(2012) uses a field experiment to measure the frictions in comparing insurance plans. Other papers have used hypothetical health choices (Johnson et al. (2013); Bhargava et al. (2015)). While our design is closest to that of Johnson et al. (2013), none of the existing papers manipulate incentives by changing the features of the plans or study the effects of incentives on effort and the interaction between incentives and education. By manipulating incentives via the premiums, our paper shows that high incentives do not affect effort, but they still matter by reducing difficulty. We also find an interaction between incentives and education on improving choices.

Bhargava et al. (2015) is the only paper we know that look directly at the effect of education on health insurance choices. Although their education treatment improves choices, this treatment is confounded by a comprehension test. Specifically, subjects who receive education are given a comprehension test before their choices while subjects who do not receive education are given the test after their choices. Besides, their experiments are not incentivized. In contrast, our experiment does not ask subjects for their understanding of the concepts before they choose a plan. We also use an incentivized setting which allows to study the interaction between education and incentives.

Moving beyond health insurance, our paper is nested within the disclosure and financial education literature. Existing papers disclose incentives by translating financial concepts (for instance, interest rate in the context of borrowing and saving) into dollar amounts (Bertrand and Morse (2011); Goda et al. (2014)). We disclose incentives by showing the dollar difference between the best option and a randomly chosen option. On financial education, there have been enough studies to prompt a meta-analysis by Fernandes et al. (2014). However, they have said little about the factors contributing the education effectiveness besides the field experiments by Drexler et al. (2014) and Carpena et al. (2017). As field experiments are limited by their ability to vary the choices, our experiment using hypothetical plans shed lights on how differences in the environment, such as premium change, can complement education.

On the broadest literature on the effects of incentives, there have been many experiments from the laboratory to the field on the effects of incentives (Camerer and Hogarth (1999), Gneezy et al. (2011)). The results are mixed and dependent on the types of tasks subjects complete. Our experiment contributes to this literature using a hypothetical choice which mimics a real-life choice and shows that incentives may matter by making choices easier.

7 Conclusion

Motivated by the poor choices in consumer finance even though stakes are high and education is freely available as documented by literature, this project implements an experiment on MTurk to study the effects of incentives and education on behaviors. We use the health insurance setting and vary stakes underlying the decisions, information about the stakes via disclosure, and access to education. To pin down the mechanisms of incentives and education, we measure both choices of insurance plans and time spent in the task.

There are three main findings of the experiment. First, high incentives have a surprising alternative channel in making the problems simpler. Second, subjects perceive education to be beneficial in either hard problems or high-stakes environment. Third, the actual return of education is positive for easy problems but zero for hard problems. The combination of the last two findings suggests that in financial choices, people could be encouraged to use education when stakes are high, but policy-makers should aim to reduce the difficulty of the problems so that the educational effort translates to better choices.

Overall, even with our full intervention of high incentives and available education, the average performance is still poor. Subjects answer fewer than half the number of questions correctly. As a result, there is much space for future research to understand choices and explore potential solutions.

References

Abaluck, J. and Gruber, J. (2011). Heterogeneity in choice inconsistencies among the elderly: Evidence from prescription drug plan choice. *American Economic Review*, 101(3):377–81.

Abaluck, J. and Gruber, J. (2016). Evolving choice inconsistencies in choice of prescription drug insurance. *American Economic Review*, 106(8):2145–84.

Agarwal, S., Skiba, P. M., and Tobacman, J. (2009). Payday loans and credit cards: New liquidity and credit scoring puzzles? *American Economic Review*, 99(2):412–17.

Andersen, S., Campbell, J. Y., Nielsen, K. M., and Ramadorai, T. (2015). Inattention and inertia in household finance: Evidence from the danish mortgage market.

Bertrand, M. and Morse, A. (2011). Information disclosure, cognitive biases, and payday borrowing. *The Journal of Finance*, 66(6):1865–1893.

Beshears, J., Choi, J. J., Laibson, D., and Madrian, B. C. (2011). *How Does Simplified Disclosure Affect Individuals' Mutual Fund Choices?*, pages 75–96. University of Chicago Press.

Bhargava, S., Loewenstein, G., and Sydnor, J. (2015). Do individuals make sensible health insurance decisions? evidence from a menu with dominated options. Technical report, National Bureau of Economic Research.

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

Caplin, A., Csaba, D., Leahy, J., and Nov, O. (2018). Rational inattention, competitive supply, and psychometrics. *NBER Working Paper*, (No. 25224).

Caplin, A. and Dean, M. (2015). Revealed preference, rational inattention, and costly information acquisition. *The American Economic Review*, 105(7):2183–2203.

Carpena, F., Cole, S., Shapiro, J., and Zia, B. (2017). The abcs of financial education: experimental evidence on attitudes, behavior, and cognitive biases. *Management Science*.

Dewan, A. and Neligh, N. (2017). Estimating information cost functions in models of rational inattention. *working paper*.

Drexler, A., Fischer, G., and Schoar, A. (2014). Keeping it simple: Financial literacy and rules of thumb. *American Economic Journal: Applied Economics*, 6(2):1–31.

Fernandes, D., Lynch Jr, J. G., and Netemeyer, R. G. (2014). Financial literacy, financial education, and downstream financial behaviors. *Management Science*, 60(8):1861–1883.

Gneezy, U., Meier, S., and Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives*, 25(4):191–210.

Goda, G. S., Manchester, C. F., and Sojourner, A. J. (2014). What will my account really be worth? experimental evidence on how retirement income projections affect saving. *Journal of Public Economics*, 119:80–92.

Gross, T. and Notowidigdo, M. J. (2011). Health insurance and the consumer bankruptcy decision: Evidence from expansions of medicaid. *Journal of Public Economics*, 95(7-8):767–778.

Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., and Bigham, J. P. (2018). A data-driven analysis of workers' earnings on amazon mechanical turk. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 449. ACM.

Johnson, E. J., Hassin, R., Baker, T., Bajger, A. T., and Treuer, G. (2013). Can consumers make affordable care affordable? the value of choice architecture. *PloS one*, 8(12):e81521.

Kaiser, T. and Menkhoff, L. (2017). *Does financial education impact financial literacy and financial behavior, and if so, when?* The World Bank.

Kling, J. R., Mullainathan, S., Shafir, E., Vermeulen, L. C., and Wrobel, M. V. (2012). Comparison friction: Experimental evidence from medicare drug plans. *The Quarterly Journal of Economics*, 127(1):199–235.

Loewenstein, G., Friedman, J. Y., McGill, B., Ahmad, S., Linck, S., Sinkula, S., Beshears, J., Choi, J. J., Kolstad, J., Laibson, D., et al. (2013). Consumers' misunderstanding of health insurance. *Journal of Health Economics*, 32(5):850–862.

Matějka, F. and McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–298.

Stango, V. and Zinman, J. (2009). What do consumers really pay on their checking and credit card accounts? explicit, implicit, and avoidable costs. *American Economic Review*, 99(2):424–29.

A Details on the Main Experiment

A.1 Subjects and Treatments

Table A.1: Number of Subjects by Treatment Arms

Treatment	Number of Subjects
LU0	253
LD0	253
HU0	251
HD0	249
LU1	249
LD1	250
HU1	251
HD1	253
Total	2009

Table A.2: Demographics: Subjects versus US population

	Subjects	US Census
Male	55%	49%
College or More	57%	31%
Race		
White	73%	60%
Black	11%	13%
Hispanic	7%	18%
Asian	7%	6%
Age (18 and more)		
18-29	30%	19%
30-39	39%	18%
40-49	16%	17%
No Insurance	18%	9%

Table A.3: Demographics Balance across Treatments

	Male	College	White	Age	No Insurance
High Incentive	-0.0124	0.00248	-0.0133	-0.0189	-0.0466**
	(0.0222)	(0.0221)	(0.0198)	(0.0546)	(0.0170)
Disclosure	0.00737	-0.0125	0.0151	-0.0292	-0.00117
	(0.0222)	(0.0221)	(0.0198)	(0.0546)	(0.0170)
Education	0.0126	0.0137	0.0181	0.0336	-0.00817
	(0.0222)	(0.0221)	(0.0198)	(0.0546)	(0.0170)
Observations	2003	2009	2009	2009	2009

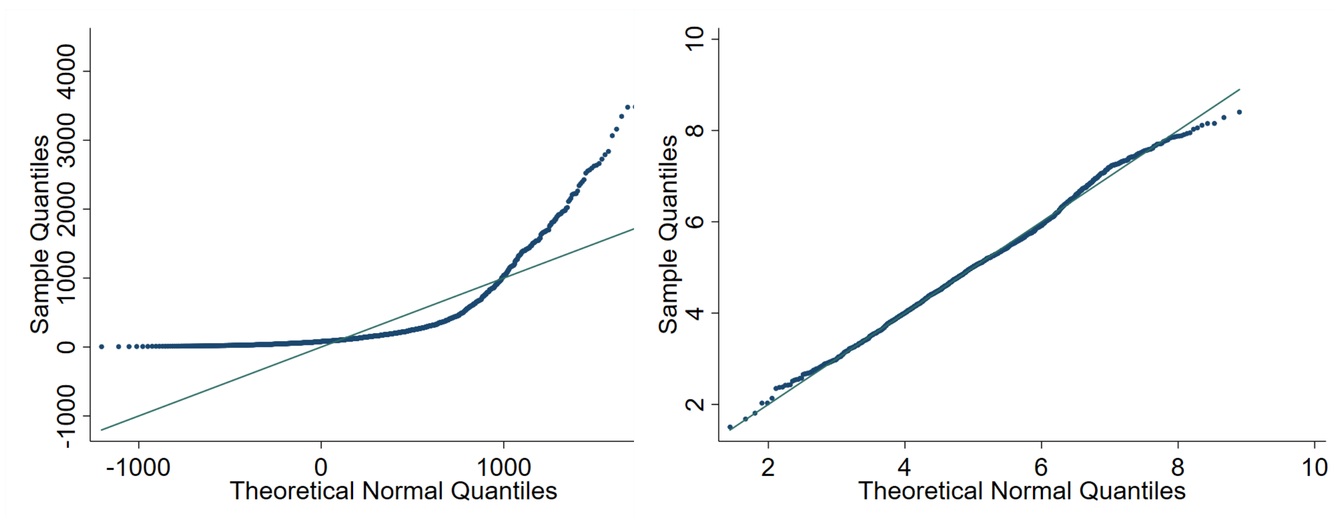
Standard errors in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

A.2 Level vs Logarithm of Time

The QQ-plots compare quantiles from a theoretical normal distribution to the quantiles generated from the level of time spent in insurance task, and those from the corresponding logarithm of time. These diagnostic plots show that the logarithm of time is closer to a normal distribution.

Figure A.1: Overall Time on Insurance Task

(a) Level of Time (seconds) (b) Level of Time (seconds)



A.3 Health Insurance Terms

Below are the glossary definitions, listed alphabetically, used in the experiment.

Coinsurance

The percentage of costs of a covered health care service you pay (20%, for example) after you've paid your deductible.

Copayment

A fixed amount ($20, for example) you pay for a health care service after you've paid your deductible.

Copayments (sometimes called "copays") can vary for different services within the same plan, like drugs, lab tests, and visits to specialists.

Deductibles

The amount you pay for health care services before your insurance plan starts to pay. With a $2,000 deductible, for example, you pay the first $2,000 of services yourself. After you pay your deductible, you usually pay only a copayment or coinsurance for covered services. Your insurance company pays the rest.

Maximum Out-of-Pocket

The most you have to pay for services in a plan year. After you spend this amount on deductibles, copayments, and coinsurance, your health plan pays 100% of the costs of covered benefits. The out-of-pocket limit doesn't include your monthly premium.

Premiums

The amount you pay for your health insurance every month. In addition to your premium, you usually have to pay other costs for your health care, including a deductible, copayments, and coinsurance

B More Results - Main Experiment

B.1 Summary

Table B.1: Summary Statistics - Insurance Task

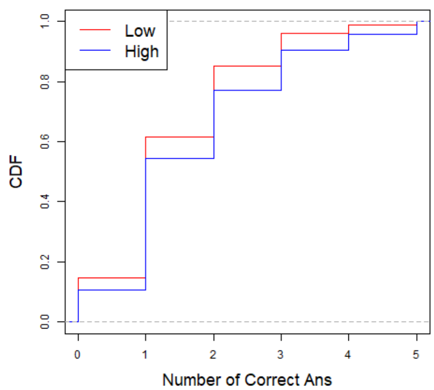
	No Edu				Edu			
	Undisclosed		Disclosed		Undisclosed		Disclosed	
	Low LU0	High HU0	Low LD0	High HD0	Low LU1	High HU1	Low LD1	High HD1
Answers	1.43	1.61	1.44	1.60	1.44	1.82	1.52	1.89
Time (sec)	296	291	296	300	392	382	362	365
EduUse (%)	.	.	.	.	28.11	32.67	35.20	33.99
EduTime (sec)	.	.	.	.	30.62	25.67	30.75	26.15

B.2 Effects of Incentives

High incentives improve performance without changing time spent, regardless of disclosure treatment.

Figure B.1: Effects of Incentives on Performance - Split by Disclosure

(a) Undisclosed



(b) Disclosed

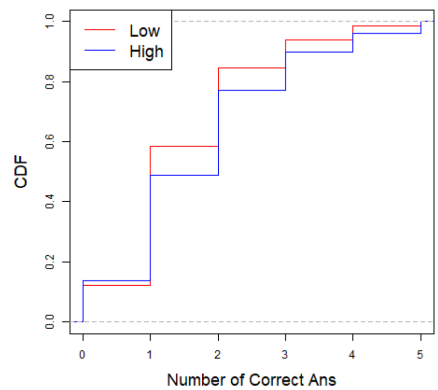
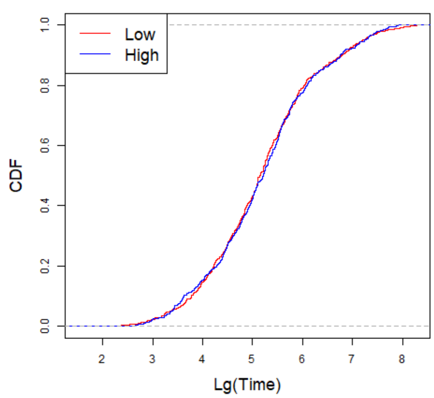


Figure B.2: Effects of Incentives on Time - Split by Disclosure

(a) Undisclosed



(b) Disclosed

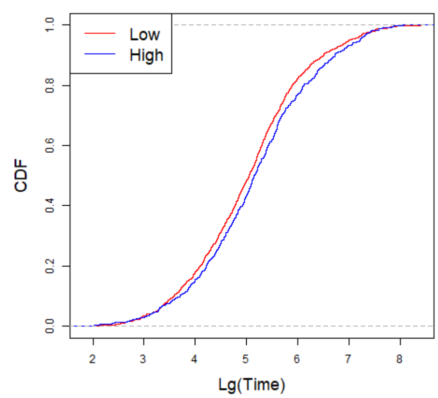


Table B.2: Incentive Perception - % of Subjects by Their Answers

	No Edu				Edu			
	Undisclosed		Disclosed		Undisclosed		Disclosed	
	Low	High	Low	High	Low	High	Low	High
$2.25	8.7	7.6	16.2	11.2	6.8	7.6	15.2	10.7
$3.5	9.9	9.2	74.3	11.6	12.4	14.3	79.2	7.9
$7	5.9	5.6	6.3	73.5	2.8	4.8	4.4	76.7
Prior: (50%: $3.5, 50%: $7)	75.5	77.7	3.2	3.6	77.9	73.3	1.2	4.7
Total	100	100	100	100	100	100	100	100

Q: If you always choose the lowest-cost plan, what do you think you earn?
Correct Answer for L: $3.50, for H: $7

B.3 Incentive and Education

Education availability improve performance only under high incentive but increase time regardless of incentive levels.

Figure B.3: Effects of Education Availability on Performance - Split by Incentive

(a) Low Incentive (b) High Incentive

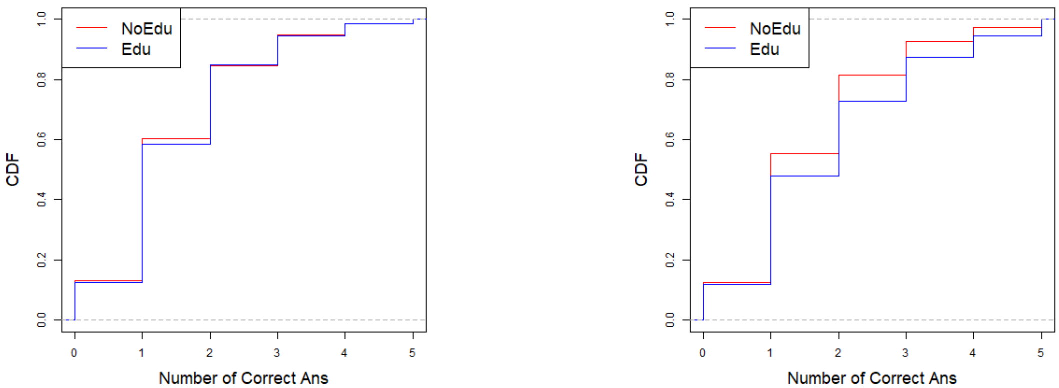
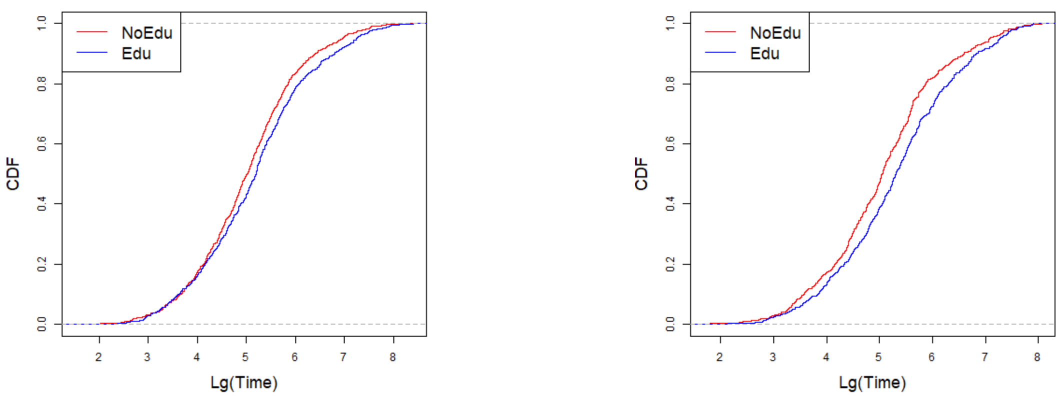


Figure B.4: Effects of Education Availability on Time - Split by Incentive

(a) Low Incentive (b) High Incentive



B.4 Heterogeneous Effects

Table B.3: Heterogeneous Effects

	Ans	Lg(Time)
θ^h	0.215**	0.605***
	(0.0908)	(0.0942)
High Incentive	0.0505	-0.120
	(0.0850)	(0.0880)
Education	-0.0618	0.0401
	(0.0827)	(0.0892)
High Incentive x θ^h	0.218	0.192
	(0.134)	(0.132)
Edu x θ^h	0.175	0.217
	(0.126)	(0.134)
High Incentive x Edu	0.273**	0.165
	(0.123)	(0.125)
High Incentive x Edu x θ^h	-0.102	-0.126
	(0.197)	(0.188)
Constant	1.311***	4.779***
	(0.0655)	(0.0666)
Observations	2009	2009

Standard errors in parentheses.

* $p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

The regressions control for the disclosure treatment.