Turning worries into performance: Results from an online experiment during Covid*

Timothée Demont[1], Daniela Horta Sáenz[1], and Eva Raiber[1]

[1] *Aix Marseille Univ, CNRS, AMSE, Marseille, France*

February 28, 2023

Abstract

Worrisome topics, such as climate change, economic crises, or the Covid-19 pandemic, are increasingly present and pervasive due to digital media and social networks. Do worries triggered by such topics affect the cognitive capacities of the youth? In an online experiment during the Covid-19 pandemic (N=1503), we test how the cognitive performance of university students responds when exposed to topics discussing (i) current mental health issues related to social restrictions or (ii) future labor market uncertainties linked to the economic contraction. Moreover, we study how such response is affected by a performance goal. We find that the labor market topic increases cognitive performance, when the latter is motivated by a goal. The positive reaction is mainly concentrated among students with larger financial and social resources, which points at an inequality-widening mechanism. Conversely, we find no effect after the mental health topic. We even find a weak negative response among those mentally vulnerable, when payout is not conditioned on reaching a goal.

Keywords: cognitive performance, financial worries, Covid-19, financial incentives, anxiety, coping behaviors

JEL codes: C91, D91, D81.

*We are grateful to Laurence Bouvard for administering the survey and to the admin team that made the project possible. We thank Michele Belot, Nicolas Berman, Björn Brey, Yannick Dupraz, Romain Ferrali, Mathieu Lefebvre, Golvine de Rochambeau, Avner Seror, Roberta Ziparo, and seminar and conference participants at Aix-Marseille School of Economics, University College Dublin, LMU Munich, FETS online seminar, EEA annual conference in Milan, and ASFEE conference in Lyon for their helpful comments. This work has been financed by the French National Research Agency Grant RA-Covid-19 LearninCov and supported by the French National Research Agency Grant ANR-17-EURE-0020 and by the Excellence Initiative of Aix-Marseille University A*MIDEX. The experiment received approval from the ethics committee of Aix-Marseille University, reference number 2020-12-03-004. Eva Raiber is a Research Affiliate at the Centre for Economic Policy Research (CEPR).

1 Introduction

Today, distressing topics regarding current problems or future uncertainties are increasingly pervasive and salient. The youth in particular is confronted with worrisome issues constantly through digital media and social networks. Several recent studies have highlighted the difficulty to resist the impulse of consuming negative news, especially during periods of crisis, with negative health consequences (e.g. Thompson et al., 2019; McLaughlin et al., 2022). Current examples include the Covid-19 pandemic, economic crises, and climate change. Do such worries affect cognitive performance? Worrying about such issues can be distracting and make it hard to concentrate on the task at hand. It can impose a cognitive burden depleting limited cognitive resources (Mullainathan & Shafir, 2013). Yet, certain types of worries might be motivating - when their effects are not yet determined and where effort can still mitigate the consequences. People with the time and means to cope with the consequences might even see those topics as a challenge. The effect on cognitive performance can also depend on what is at stake, e.g. is there an explicit goal that one can focus on?

In an online experiment, we exogenously vary the type of worry that participants have on their minds. During the COVID-19 pandemic, we confronted university students with two real-life concerns - (i) current mental health issues related to social restrictions, or (ii) future labor market uncertainties - and compare their cognitive performance to a control group that was shown a neutral topic. The treatments are motivated by the evidence that during the pandemic young people worried most about uncertain employment opportunities and that social restrictions burden their psychological well-being (OECD, 2020; Fetzer et al., 2020; Pierce et al., 2020; McGinty et al., 2020; Etheridge & Spantig, 2022; Santé publique France, 2021). We then vary how performance is rewarded. We either pay participants for each correct task or condition payment on reaching a goal. Finally, we explore heterogeneity in the response focusing on characteristics linked to financial or mental health vulnerability that can decrease the ability to cope with the consequences of the pandemic.

We ran the experiment with around 1500 students at Aix-Marseille University, a French public university with a mixed student body, between February and April 2021. The treatment and control topics included a newspaper article, two graphics, and comprehension and reflection questions. The

mental health topic discussed the situation of students during a lockdown. France was just out of its second national lockdown and partial lockdowns were progressively put in place at the regional level. The topic thus increased the salience of a current issue that participants had been unwillingly confronted with. The labor market topic discussed the projected labor market situation of young adults. With students not being in the labor market yet, the topic could stimulate worries about an uncertain future. Yet, the topic has "scope for action" as students might still be able to influence the outcome through increased effort.

We measure cognitive performance with a Raven-matrices-like task. Students were either faced with a standard gradual payment scheme for each correct answer (piece-rate treatment) or received this payment only upon reaching an achievable minimum level (threshold treatment). The threshold treatment thus had an assigned goal that can result in a "fail" or "pass" situation – similar to an exam environment.

After the treatment topics, students state feeling worse and more nervous than after one of the control topics. Yet, we find that under the threshold payment scheme, participants assigned to the labor market topic increase their cognitive performance by 7%, relative to the control group. At the same time, the mental health topic has no such effect. With piece-rate payment, both topics have no significant effect on cognitive performance. However, those that are mentally vulnerable, with a high depression score, have weakly lower cognitive performance when assigned to the mental health topic. We further explore heterogeneity using a causal machine-learning method (Causal Forest (Athey & Wager, 2019)). We find the positive treatment effect of the labor market topic under the threshold payment scheme to be concentrated among those in a financially stable position and a well-off family background, those with a more active social life during the pandemic, and those further away from graduation.

We investigate possible mechanisms of the positive effect of the labor market topic under the threshold payment scheme. The labor market topic is the most likely to trigger a motivating effect. It puts the focus on students' chances in the labor market – an issue that students are not necessarily faced with in the near future and that they can see as a challenge to overcome. This interpretation is underlined by the observation that the positive effect is driven by those that (i) still have time to influence their labor market outcome through effort, and (ii) have more financial and social resources to cope with negative outcomes. We also find that participants rate a "successful career"

and "good grades" as more important after the labor market topic. This can interact with the threshold payment scheme that puts the focus on performing well to reach a specific goal. Indeed, we find that the treatment effect is the strongest among those that are close to the threshold before treatment. On the contrary, those that are financially vulnerable are not benefiting from the motivational boost. Thus, financial worries in this setup have a potentially inequality-increasing effect.

Our experiment relates to the ideas in the psychology of poverty literature on the relationship between cognitive functioning and scarce financial resources. Scarcity theory suggests that worries – e.g., poverty, financial strain, uncertainty – impose a mental burden that taxes cognitive function and alters how people process information and make decisions (Shah et al., 2012; Haushofer & Fehr, 2014; Schilbach et al., 2016). Mani et al. (2013) find that making financially constrained people think about a worrisome financial decision decreases cognitive performance. However, this could not be replicated in other contexts (González-Arango et al., 2021; Dalton et al., 2020).[1] Contrary to the existing experiments in the literature (e.g. Mani et al., 2020; Dalton et al., 2020), we do not ask participants to contemplate a decision while they go through the cognitive performance task. Rather, we postulate that highlighting real-life worries can lead to inattention or have an effect on motivation.[2] Moreover, we are the first to test if varying the incentive structure for cognitive performance interacts with worries.

The literature suggests that financial vulnerability can weaken resilience to financial stress. For instance, Duquennois (2022) finds that students from a lower socio-economic status background score worse on mathematical exam questions that make large sums of money salient. She suggests that financial salience can capture the attention of those financially vulnerable.[3]. We run our experiment in a public university, accessible to a large share of the population. 45% the participants qualify for the means-tested state scholarship[4] and 28% state that they are struggling financially.

[1]The mixed and often non-significant results led some researchers to argue recently that economic rationality might be unaffected by temporary impairments in cognitive resources (Canavari et al., 2019; Achtziger et al., 2020).

[2]The null results in the piece-rate treatment suggest that at least on average participants were able to focus on the cognitive performance task independent of the topic. This could be due to the relatively high incentives for the task (1€ per correct answer, allowing them to earn up to 10€ in 4 minutes) or the topics not being sufficiently distracting. However, our results suggest caution as there might be an effect on the most vulnerable, such as those with a high depression score.

[3]Related, Carvalho et al. (2016) find no changes in cognitive performance before and after pay-day - a result, however, that is reexamined by Mani et al. (2020) and investigated further by Farbmacher et al. (2021). Finally, Kaur et al. (2021) find that workers are more productive after they receive their cash payments.

[4]Compared to a nationwide scholarship rate of 38% in 2020-2021(see https://www.enseignementsup-recherche

Our sample thus includes a large share of young people that are financially vulnerable allowing us to investigate heterogeneity along these lines. Our results that those with larger financial resources might benefit from troubling topics and those vulnerable do not, or even perform worse, is in line with the previous results in the literature.

We also contribute to the literature by documenting the consequences of economic and social anxiety in the wake of a global pandemic. It adds more broadly to the behavioral economics literature about the impact of stressful events, economic uncertainty, and mental load on cognitive performance and economic decisions (see Deck & Jahedi (2015) and Schilbach et al. (2016) for recent reviews). Closely related, Bogliacino et al. (2021) find that people affected by negative Covid-19 shocks perform worse in a cognitive reflection task and are more risk-loving. However, they find that reminding participants of negative emotions does not affect their performance. Our results suggest that while students state feeling worse after the worrisome topics, these negative emotions – at least for most – do not decrease cognitive performance.

Our study draws upon the literature on financial incentives, effort and task performance (see Bonner & Sprinkle (2002) for a review). Various laboratory studies have shown that the financial incentives structure can affect task performance (Wright, 1990; Earley & Lituchy, 1991; Jenkins Jr et al., 1998). Yet, others found that financial incentives, while inducing more effort, do not necessarily improve performance (See e.g., Cole et al. (2018), Gignac (2018), Belle & Cantarelli (2015), Camerer & Hogarth (1999)) and can even decrease it (e.g., Hickman & Metz (2015), Ariely et al. (2009), Baumeister (1984)). Indeed, monetary incentives can trigger an emotional response that counteracts the motivating effect for some individuals, especially those with low expectations (Castro et al. (2021)).

Among the different incentive schemes, the threshold-based schemes seem to have the most consistent positive effect on effort. They make both the goal and the link between pay and performance explicit and salient (see Bonner et al. (2000)). From a theoretical point of view, it should motivate those who are or believe they are close to the threshold.[5] However, threshold-based payment schemes could have a demotivating effect among those believing to score below the threshold – either because they are not incentivized or by causing negative emotions lowering their motivation

.gouv.fr/sites/default/files/2021-09/nf-sies-2021-20-12998.pdf).

[5]For example, Levitt et al. (2016) evaluate the financial incentives for meeting an achievement standard in high school and find that the program has a large effect on students at the threshold.

and effort. We add to this literature by showing how circumstances - being confronted with worrisome thoughts – and participants' characteristics can change the response to different incentive schemes.

2 Experimental Design

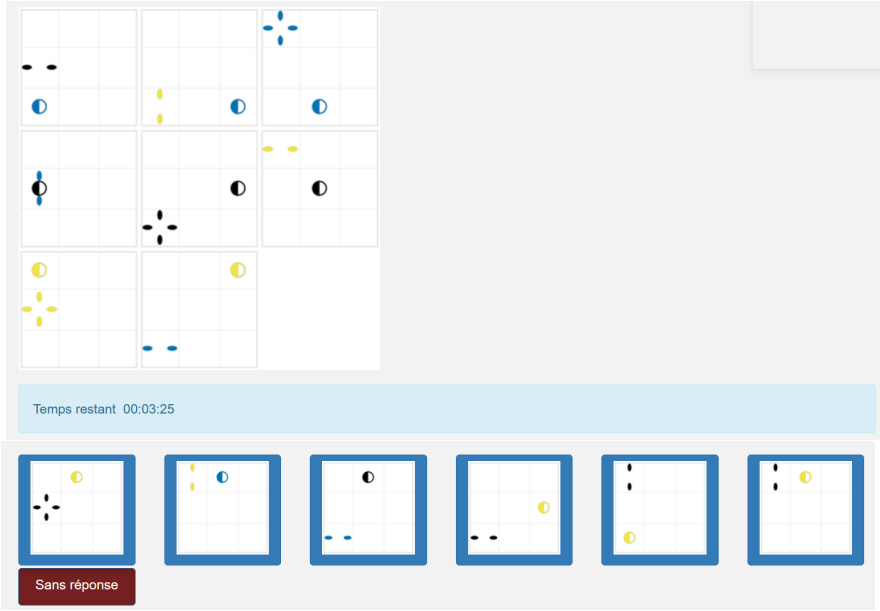
2.1 Recruitment

Participants were recruited from Aix-Marseille University (AMU), the largest French-speaking university with around 80 000 students. Interested students were invited to sign up for a paid online survey, approved by the AMU ethics committee, by mail. Students signed up with their unique official university email addresses. Between February and April 2021, for six weeks, 500 students who had signed up were randomly selected and sent an individual survey link on a Tuesday that was valid until the Friday of the same week. Participants had 90 minutes to finish the survey once started. Participants received payment for completing the survey of 7€ paid in the form of a voucher. Out of the 500 students invited each week, on average, 52% started it. For the last week of invites, all those previously not selected received the survey link, as well as those who had been invited but had not completed the survey.

2.2 The Covid context

At the time of the survey, France was experiencing another Covid wave, reaching 39,000 daily cases at the end of March for the first time since November 2020. The country was not in lockdown, as the second national lockdown had ended on December 15. However, there were important restrictions in place, such as a curfew, closure of restaurants, cultural institutions and large shopping centers, severe international travel restrictions, and remote or hybrid organization of classes. During March, partial lockdowns were progressively reintroduced at the regional level, though not in the region of this study. Importantly, while elderly people were starting to get vaccinated, vaccination was not yet available for students (it would not be until mid-May).

Figure 1: Example of an incomplete matrix to measure cognitive performance.



Note: Matrices taken from the matrix reasoning item bank of Chierchia et al. (2019).

2.3 Main outcomes

Cognitive performance: Cognitive performance, sometimes referred to as cognitive ability or fluid intelligence, is defined as the capacity for reasoning, problem-solving, and abstract thinking. It involves the ability to process information quickly and to go beyond memorization or imitation (Diamond, 2013).

To measure cognitive performance we use matrices from the matrix reasoning item bank of Chierchia et al. (2019), a collection of open-access abstract reasoning items, similar to the Raven's Matrices (Raven et al., 2003). Participants were shown an incomplete matrix containing colorful, abstract forms with one missing field and were asked to select the missing item among six options. Figure 1 shows an example.

In the first round - the training task -, participants were shown one example and then asked to correctly solve 4 matrices. They had a time limit of 3 minutes (45 seconds per matrix). For each correctly solved matrix, they received 0.5€. We use the number of correctly solved matrices of this first round as "baseline cognitive performance score". In the second round which took place after the treatment, participants were asked to correctly solve 10 items with a time limit of 6 minutes and 40 seconds (40 seconds per matrix). The payment scheme for the second round varied by

treatment. The number of correctly solved matrices in the second round is our main outcome of cognitive performance.

Emotional State: Participants were asked a short translated version of the multidimensional mood questionnaire (MDMQ, Hinz et al., 2012) to measure their current emotional state. This version of the MDMQ consists of 12 questions along three dimensions: feeling good versus bad, feeling awake versus tired, and feeling calm versus nervous. For each mood dimension, four questions are asked, two phrased positively and two negatively. Importantly, the MDMQ explicitly asks how the respondent feels at this current moment. Half of the participants were asked about their emotional state before the treatment and half after the treatment but before the incentivized tasks.[6] This was cross-randomized with the topic treatments and the payment schemes.

2.4 Other tasks and measures

Cognitive reasoning: We used three questions in the style of Frederick (2005). In each case, what seemed to be the intuitive answer was not the correct answer. Participants had 4 minutes and 30 seconds to answer the questions (1.5 minutes per question). Participants could earn up to 3€ in this section (1€ per question).

Risk-taking: To measure risk-taking, we used a lottery choice in the style of Gneezy & Potters (1997). Participants could invest up to 3€ from their baseline payment. They had a 50% chance to triple their investment, and a 50% chance to lose their investment.

Revealed preferences for a coaching program: Participants were offered to participate in a lottery to win an individual online coaching program of a market value of 385€. The coaching program included a test and three individual sessions with a coach. The program was proposed by one of the leading companies in this sector with our cooperation. Participants could choose between different modules: interview simulation, work methodology, self-confidence and stress management, and psychological support.

Depression and Anxiety score: We compute a Depression score through the Patient Health Questionnaire-9 (PHQ-9, Kroenke & Spitzer, 2002)[7], and an Anxiety score through a short version

[6] We administered the MDMQ only to half of the sample after the treatment as we were worried that the time lag between the treatment and the outcome measures might play a role. However, we do not find any effect of answering the MDMQ after the treatment rather than before on any outcome.

[7] We changed the last question from the PHQ-9, which explicitly asked for the presence of suicidal thoughts, to one related to depression from HADS.

of the Hospital Anxiety and Depression Score (HADS, Zigmond & Snaith, 1983).[8] Both are standard mental health self-administered questionnaires inquiring about the intensity of depressive and anxiety symptoms in the last few weeks: the higher score, the worse the depression and anxiety symptoms.

Locus of control: We used a short version of The Internal Locus of Control Index (ICI, Duttweiler, 1984). This index measures to what extent subjects feel they have control over their lives. Highly internal subjects feel responsible for the things that happen in their lives, while low internal subjects believe that factors beyond their control determine their lives.

2.5 Treatments

In the experiment, we cross-randomize the reflection topic and the payment scheme.

Topic treatments: Participants were randomly shown one of four topics and asked to reflect on them. Each topic contained an article of around 600 words including two graphical illustrations followed by some non-incentivized comprehension questions. The topics also included several reflective questions to motivate the student to think about the topic and their situation. The format, length and number of questions were the same for all topics.

Both the Labor Market and Mental Health topics included information on the negative consequences of the Covid pandemic and the lockdowns. For the control groups, we chose two different topics: one article about the progressive elimination of cage rearing in France (Animal Welfare) and one article on the future of the Artemis program to land humans on the Moon again (Space Program).

All articles were taken from online platforms of actual newspapers and reflect information that students are confronted with daily. While addressing negative topics, we purposely chose articles that were factual and not sensational.[9] The treatments were designed to make the labor market or mental health consequences of the Covid pandemic salient and to have participants reflect on their situation. The articles were included to give context and substance to the treatment but generally not to provide novel information about the Covid consequences. Indeed, only 16% of participants stated that they "learned a lot" from the labor market topic, and 9% said the same from the mental

[8] For both measures included the option not to answer the question.

[9] As we mention later, at the end of the survey, we provided links to the university's and general support programs. Those that did not finish the survey received a follow-up mail with a condensed version of this information.

health article. The controls were designed to occupy the participants for the same time and to also put them into a "reflective state" without worrying or stressing them.

Labor Market topic:

The labor market topic started with an article about the difficulty of young graduates entering the labor market. It mentioned the decreased number of job offers due to the pandemic and described the expected increase in unemployment. It included two graphs, one illustrating the expected increase in unemployment, and one highlighting the pessimistic view that many young people have about their labor market prospects. The reflective questions asked about the participant's views on their labor market perspectives and their economic situation.

For most students, entry into the labor market is not an immediate issue. Furthermore, their perspectives can still be affected by their current effort in their studies. As labor market perspectives differ by the level and field of study, we investigate heterogeneous treatment effects in this regard. It could also affect those from a lower socio-economic background more than those with affluent parents.

Mental Health topic:

The mental health topic included an article about the psychological effects of the pandemic, focusing on the isolation of young people due to national lockdowns. It included a graph that illustrated the depression rate for different age groups and a graph displaying how prevalent mental health problems, stress and anxiety are. The reflective questions asked about the participants' stress, feeling of isolation and regret about their social life.

As explained above, although France was not in lockdown at the time of the survey, there were still heavy restrictions in place, especially affecting students (e.g. remote or hybrid classes, a curfew, closed bars and cultural institutions). The article thus spoke about a current issue that students could not do anything against. While it surely resonated with all students, some are potentially more vulnerable - in particular those with a worse baseline mental health.

Payment schemes: For the second round of cognitive performance, participants were randomly allocated to one of two payment structures, cross-randomized with the topic treatments:

Piece-rate payment: In the piece-rate treatment, participants received 1€ per correctly solved matrix.

Threshold payment: In the threshold treatment, participants received 1€ per correctly solved matrix only if they correctly solved at least 5 matrices. If they solved less than 5 matrices, their payout was 0€. If they solved 5 or more, their payout was the same as in the piece-rate treatment. The payoff structure was illustrated in a table.

The threshold payment scheme can be motivating as it gives participants a goal to focus on. This should be especially relevant for those close to attaining the threshold. However, it can also add another source of stress which could lower performance. Also, if the threshold is too high to be judged as achievable by the majority of participants, it can lead to them making no effort in the task. We set an achievable threshold at solving half of the matrices correctly: A threshold that is reached by 78% of the participants in the control treatment under the piece-rate scheme.

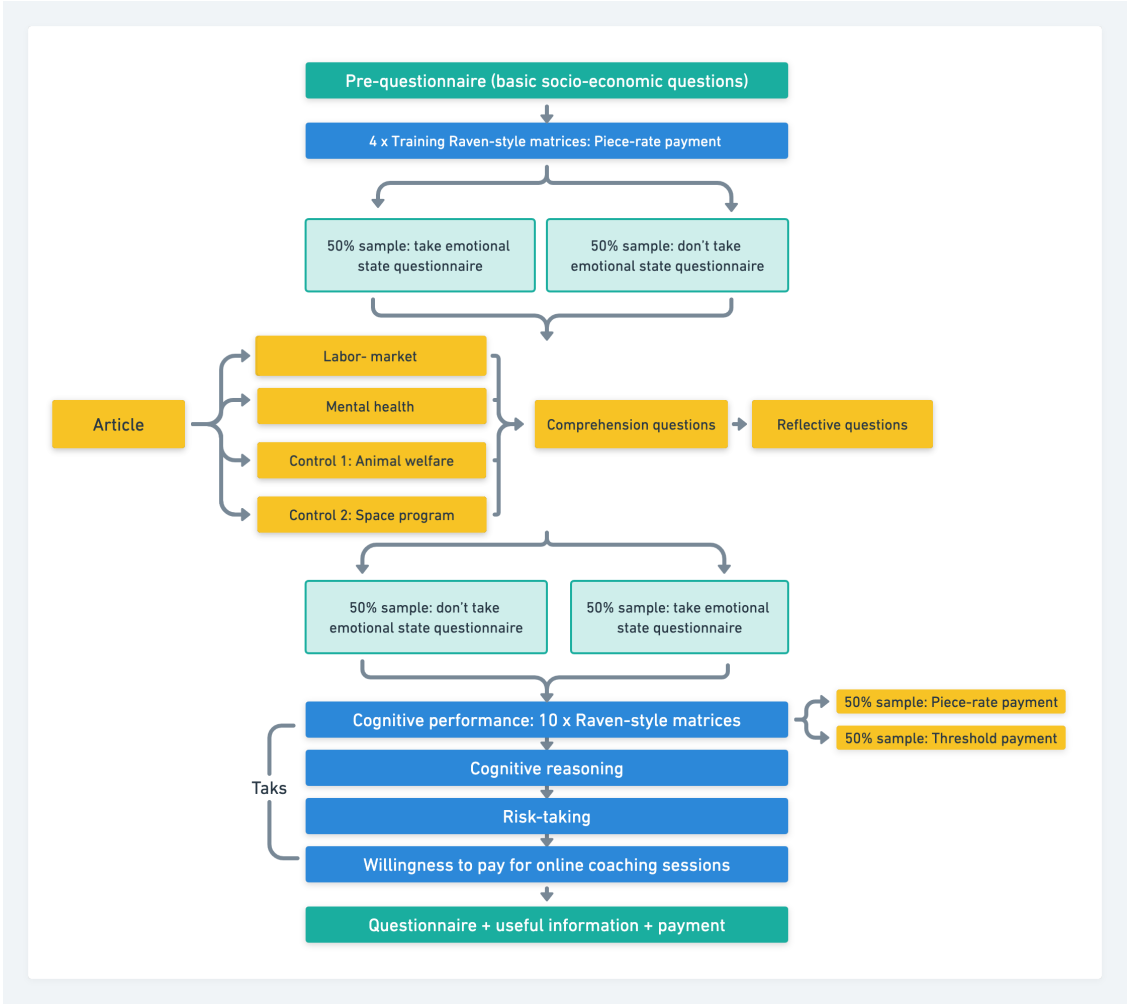
2.6 Survey structure

Figure 2 summarizes the survey structure. The survey started with an information and consent page which described the survey structure and the topics covered. The students were told that the survey would cover topics related to the pandemic. Following, respondents were asked some basic socio-demographic questions in the pre-questionnaire (age, scholarship recipient, field of study, gender). All participants then faced the first round of the cognitive performance task, incentivized by a linear payment scheme. Before the treatment articles, half of the sample randomly selected were asked questions about their current mood. The other half were asked the same questions after the treatment section. Participants were then moved on to the treatment topics.

After the treatment topics and the questions about their current mood, participants were faced with the second round of the cognitive performance task with the different payment structures. This was followed by incentivized measures for cognitive reasoning, risk-taking, and the willingness to pay for an individual online coaching program.

Following, respondents were asked questions about their study and career expectations and pressures, their Covid and lockdown experience, their current social habits, and their financial situation. This part included the questions for the additional measures (mental health, anxiety, locus of control). The survey ended with a questionnaire on the socio-demographics of the student and their family. In the end, participants were informed of their payment and could choose the method of payment (Amazon or Cultura voucher).

Figure 2: Survey structure



2.7 Ethics

In the experiment, students were purposely faced with reflection topics that can be expected to trigger negative emotions. To minimize the risk of an effect that extends beyond the duration of the experiment, the following steps were undertaken. First, students were informed that the survey would deal with the pandemic when they signed up and when they started the survey. Students could end the anonymous survey at any moment. Second, the informational material in the reflection topics (article, graphics), though negatively framed, were taken from standard newspapers and official organisations and judged "non-sensational". They thus reflect information in a format that young people are constantly confronted with, presumably multiple times a day. The reflection questions were questions that young people are generally faced with as well. Third, at the end of the survey, participants were provided with additional information about the university's support system and other relevant Covid information if they were interested. Participants who did not complete the survey after reaching the topic treatment stage were sent an email to inform them about the university's support system. This procedure received approval from the ethics committee of Aix-Marseille University, reference number 2020-12-03-004.

3 Data

3.1 Descriptive Statistics

Of the 1562 students that started the questionnaire, 1503 students finished it.[10] The overall rate of attrition is thus 3.8% with no differential attrition between the treatment groups. 779 participants played under the piece rate and 724 under the threshold payment scheme. 352 participants saw the labor market, 359 the mental health, 386 the animal welfare, and 406 the space topic.

Table A.1 reports the average characteristics, including the mean differences and p-values of treatment and control groups, for pre-registered baseline variables. 66% of the respondents are female, the average age is 21.6 years. 70% are in their undergraduate studies, 27% in their masters or equivalent, and 3% are doing a PhD. 37% are from Science and Technology, 25% of the students

[10] As specified in the pre-analysis plan, we excluded respondents that took less than 8 minutes (20% of the median time) or more than 100 minutes to respond to the survey, as well as participants younger than 18 years and older than 30 years.

are within the field of Law, Economics and Management, followed by Humanities and Social Sciences (17%), Art and Languages (14%) and Health Science (11%). Around 45% receive the means-tested state scholarship that is paid when the parents have less income than a given threshold.

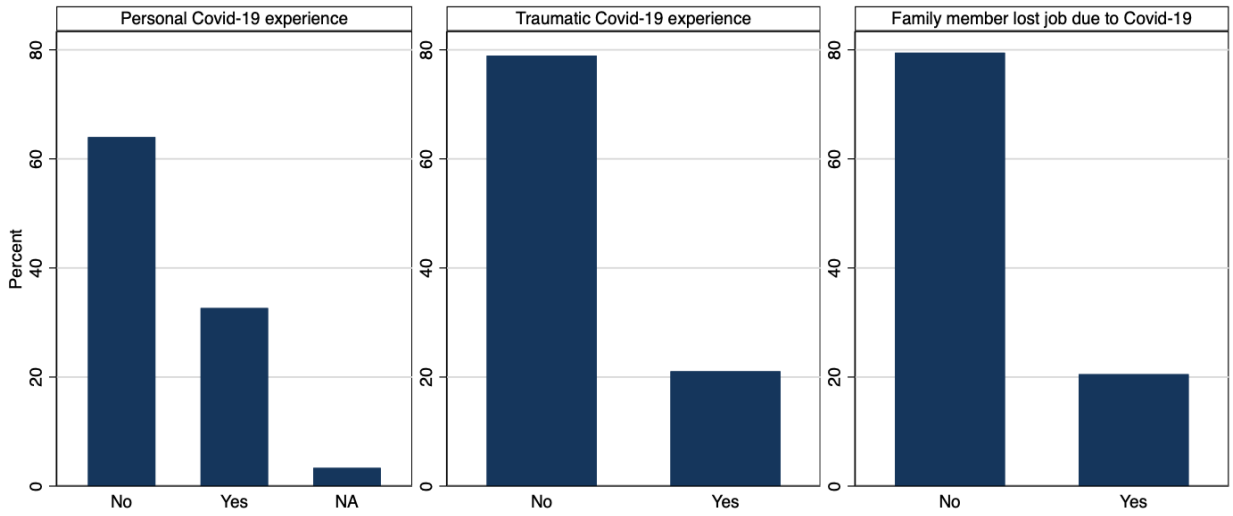
The distribution of the covariates is balanced across the treatments on the pre-registered covariates (see Table A.1). Table A.2 and A.3 show the covariates separately for the two payment structures which are also jointly insignificant. The joint orthogonality test is insignificant when comparing the topic treatments and the payment structures. Yet, there are some variables that show up significant in some specifications. As pre-registered, we include these covariates as baseline control in all our specifications.

3.2 Covid-19 experience

The experiment was run during the Covid-19 pandemic and uses Covid-related topics. We thus gather information about the individual Covid experience up to April 2021. We also summarize answers to the reflective questions asked in the treatment topics about the economic and social consequences of the pandemic.

We observe that a large proportion of students have had, one way or another, a close experience with Covid-19 by April 2021 (see Figure 3). One in three students reported having had a personal Covid-19 experience, i.e., an immediate family member, housemates, or themselves got the virus. About 20% declared having had a hospitalized family member or known personally someone who died of Covid-19. Furthermore, in our sample, 20% of students reported having someone in their family who has lost their job due to Covid-19.

Figure 3: COVID-19 experience.



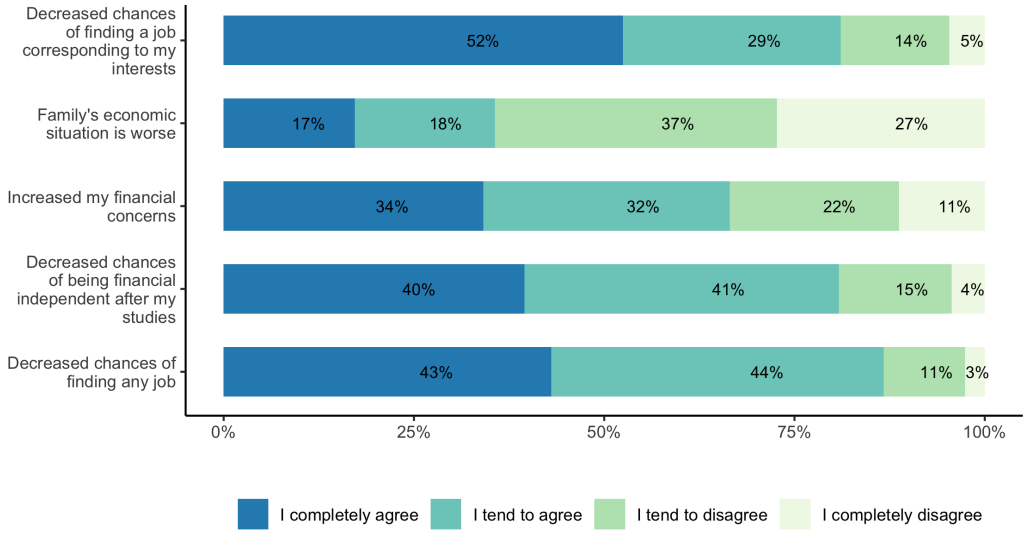
Note: Illustrates the percentage of students closely affected by COVID-19 by April 2022 in the following ways: Personal experience, if an immediate family member, housemates, or themselves got the virus. Traumatic experience, if they had a family member in the hospital or knew someone who died of Covid. Lastly, if they had a family member that lost their job due to Covid.

Figure 4 Panel A displays the perceived economic consequences of Covid-19 which we gather from the participants in the labor market treatment. We observe that the economic consequences of the pandemic are indeed a worrisome topic for students. It not only affects their economic situation and job opportunities but also undermines their career projections. 81% of students agree (completely or tend to agree) that the Covid-19 crisis decreases their chances of finding a job that corresponds to their interest. About 80% of students state a decrease in the chances of being financially independent after their studies. At the same time, students perceive a worsening in their family's economic situation (35% agreeing) and an increase in their financial concerns (36% agreeing).

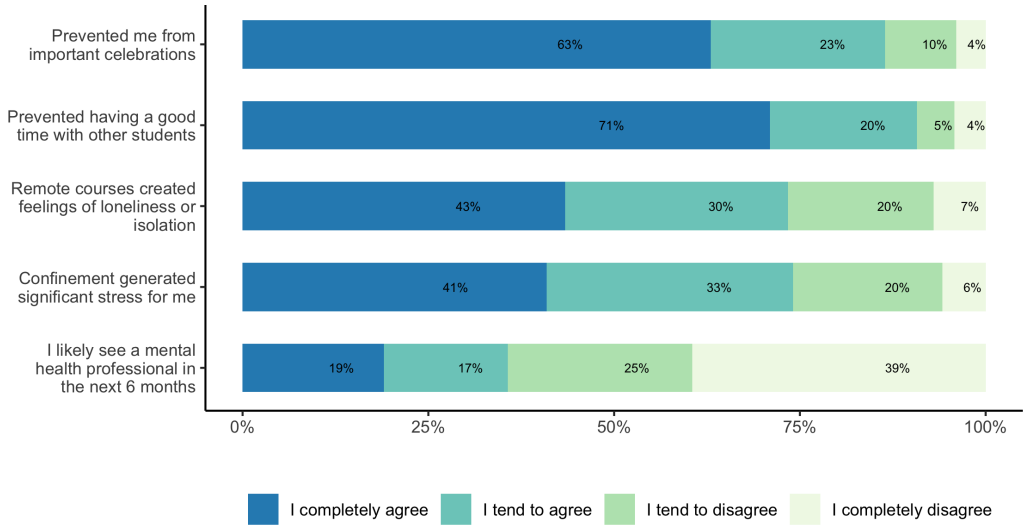
Likewise, participants indicate how they perceive the social consequences of the Covid-19 crisis in the mental health treatment (see Figure 4 Panel B). Most students agree that the health crisis has prevented them from important celebrations and good times with friends. 48% of students strongly agree that online classes create feelings of loneliness or isolation, and a further 28% tend to agree with it. Most students state that the lockdown generates a significant source of stress with 76% agreeing or tending to agree. Concerning seeking professional mental health help, 36% of students state that they will likely talk to a mental health professional in the next six months.

Figure 4: Perceived economic and social consequences of COVID-19 (February - April 2021)

a) Economic consequences of Covid-19.



b) Social consequences of Covid-19



Note: This figure plots the perceived economic and social consequences of the pandemic for those students who were assigned to the labor market (Panel A) or mental health topic treatment (Panel B). Panel A includes responses from students in the labor market treatment (n=352). Panel B includes responses from students in the mental health treatment(n=359).

4 Results

4.1 Emotional state

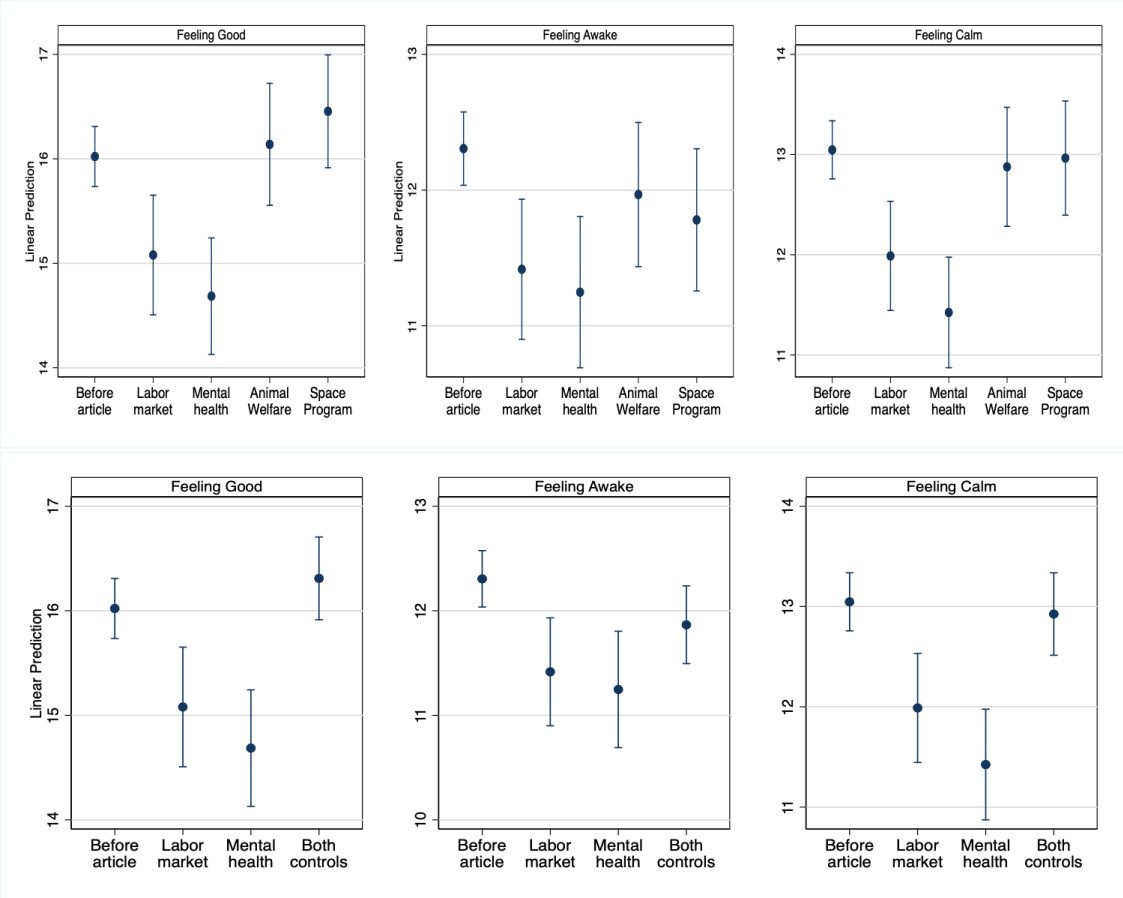
We first verify if the treatment topics had an effect on the current emotional status of the participants. Figure 5 illustrates the responses to the mood questionnaire for the three mood dimensions

before the treatments and afterwards. The specifications control for the pre-registered baseline controls (gender, age, field of study, year of study, number of correct matrices in the first round and state scholarship recipient). The regression results are displayed in Appendix A Table A.4.

We find that the two control topics do not affect the participants' emotional state significantly on any dimension compared to those asked before, except for a weak effect on making respondents more tired (significant only at 10 percent). Notably, the emotional states of participants after either of the control topics are nearly identical. Therefore, we pool the two control groups to compare them to the two treatment groups.

We find that participants state feeling significantly worse after facing either of the two treatment Covid topics compared to the control topics and compared to those asked before the topics. Participants are also significantly more tired after reading the mental health article and weakly so after reading the labor market article. Finally, they are significantly less calm after reading the two treatment articles compared to before the article and compared to those reading the control articles. Overall, the treatments had a negative effect on the emotional state of the participants, whereas the control treatments had the same, non-significant effect. Yet, we cannot exclude an experimenter demand effect as participants could feel that after the treatment topics they are expected to feel worse.

Figure 5: Emotional states before reading the treatments and afterwards



Note: Illustrates the linear prediction of emotional state before and after the topic treatment. Scores are based on four questions for each mood (two positively phrased, two negatively). Minimum possible 4, maximum possible 24. Includes pre-registered baseline controls: Gender, age, field of study, year of study, number of correct matrices in the first round, scholarship recipient (see Table A.4). N = 1503.

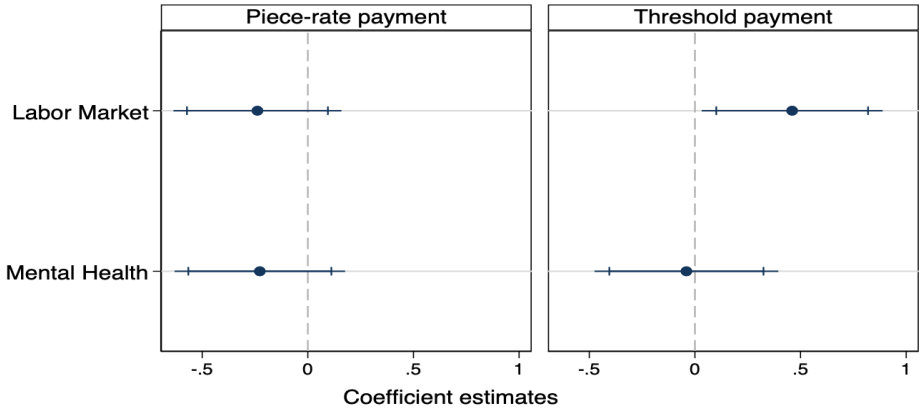
4.2 Cognitive performance

We now investigate the effect of highlighting a Covid-related topic on cognitive performance. We differentiate between the two payment structures: the piece-rate payment and the threshold payment. Figure 6 illustrates the treatment effects on cognitive performance, measured through the number of correct matrices, for each of the payment schemes. Under the piece-rate payment, the coefficient for both treatments is negative but not significantly different from zero. However, under the threshold payment, the labor market treatment positively and significantly affects students' cognitive ability. Treated students improved their performance by 7% relative to the control group mean. In contrast, the mental health treatment does not lead to any significant effect. Table A.5

summarizes the results.

We obtain similar results when the two treatments interact with the variation in the payment scheme. Interestingly, the payment scheme alone does not have a significant effect on cognitive performance. It is the combination of the threshold payment and labor market treatment that enhances the cognitive performance of students.Table A.6 displays the results.

Figure 6: Treatment effect on cognitive performance



Note: Illustrates the treatment effects on cognitive performance according to the payment scheme. The dependent variable is the number of correct matrices. Minimum possible 0, maximum possible 10. All specifications include pre-registered baseline controls: Gender, age, field of study, year of study, number of correct matrices in the first round, and scholarship recipient. Displays 90% and 95% confidence intervals. N = 1503. Based on the results displayed in Table A.5.

4.3 Heterogeneous effects

Pre-registered heterogeneity: We pre-registered heterogeneity for gender, receiving a state scholarship as a measure of parental income, the field of study, the level of study, being close to finishing their studies, depression and anxiety score, and if the mood questionnaire was asked before or after the treatment topics. Figure 7 illustrates the treatment effect for the different subgroups for the piece-rate treatment (Panel A) and the threshold payment (Panel B).

Under piece-rate payment, we find no treatment effect of the labor market topic in any pre-registered dimension, except for among students in "health science" from which the sample is too small to draw conclusions. For the mental health topic, we find that it decreases cognitive perfor-

mance among those without a scholarship and those with a depression score above the median.[11] We expected those with a scholarship to be more vulnerable as they come from a poorer background. Yet, receiving a state scholarship might also give students a stable financial income and thus improve their financial stability compared to those who do not receive it but have a similar financial background. The effect on those with poor mental health is in line with expectations: Those that are especially vulnerable perform worse. The coefficient for the anxiety score goes in the same direction, though is not significant. Finally, we verify that the emotion questionnaire does not generate any treatment effect by itself.

Under the threshold payment, the treatment effect of the labor market topic is consistently positive for all subgroups. The treatment effect seems to be driven by women and those not just before labor market entry (though none of the coefficients are significantly different from each other). The mental health topic does not have an effect on cognitive performance for any pre-registered subgroup.

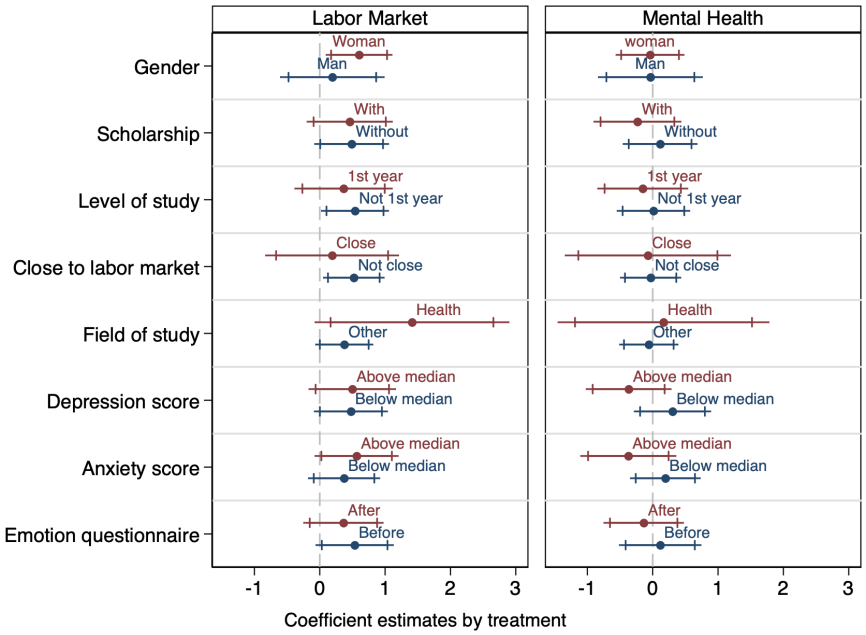
Table A.7 summarizes the heterogeneity analysis for the piece-rate treatment and Table A.8 for the threshold treatment. Table A.9 summarizes the analysis with interaction terms and tests for a general treatment effect of the threshold payment.

[11] We find that the mental health topic negatively affects the depression score though the questions were asked about the previous weeks (see Table A.14 column 1). We verify if assignment to treatment changes the group composition into those below and above the median. If treatment changed the composition, the subgroups would not be comparable between treatments. However, we find that the median for each treatment cell is the same, such that using the overall median to divide the participants into two groups and the group-specific median leads to the same results. Results are also the same if we correct the mental health score by the treatment effect (see Table A.11). We also find in a quantile regression that the mental health topic negatively affects those with an already high score. Table A.10 summarizes the results.

Figure 7: Pre-registered heterogeneity

a) Piece-rate payment scheme

b) Threshold payment scheme



Note: Panel A illustrates the differential treatment effect on cognitive performance for the piece-rate payment and Panel B for the threshold payment (see Table A.7 and Table A.9) for the pre-registered groups. The dependent variable is the number of correct matrices. Minimum possible 0, maximum possible 10. All specifications include pre-registered baseline controls: Gender, age, field of study, year of study, number of correct matrices in the first round, and scholarship recipient. 90% and 95% confidence intervals. "Emotion qnt" refers to the emotions questionnaire being asked before or after the topic treatment. N=1503

Causal Machine Learning: Following Wager & Athey (2018), we use a "Causal Forest" to uncover subgroups that react differently to our treatments in a data-driven approach. This heterogeneity analysis allows us to go beyond the pre-defined subgroup analysis by accounting for high dimensional combinations of covariates that would be hard to model otherwise.

To do so, we estimate the Conditional Average Treatment Effect (CATE) on a vector of observable characteristics, including baseline controls and a large number of covariates that provide information on participants' financial situation, expectations, family background, mental health measures, Covid-19 experience, and some self-perception questions. We then use the predicted CATE to rank the observations from those with the lowest CATE to the highest CATE and group them into quartiles. Next, we study the differences between each group. Appendix B describes the methodology in detail.

We apply the causal forest to each of our treatments. However, after assessing the quality of the forest's estimates,[12] we only detect heterogeneity in the labor market treatment and the threshold payment — the only treatment where we find an average treatment effect.

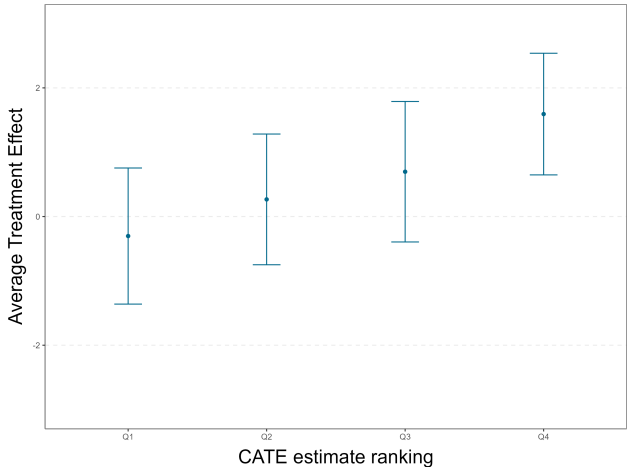
Therefore, we limit our analysis to this treatment arm. Figure 8 displays the Average Treatment Effect (ATE) on cognitive performance within each group: As expected the ATE increases monotonically along with the groups – students that belong to the first quartile have a lower treatment effect than those in the second quartile, and so on. Those in the fourth quartile are those who benefit from the treatment and experience an increase in cognitive performance. To compare the two most contrasting groups, we analyze the difference between those in the first and the fourth quartile to uncover the characteristics that predict whether or not respondents benefit from the treatment.[13] Table 1 displays the results.

Similar to the pre-register heterogeneity analysis, we find that younger students, students without a scholarship, and students not close to graduation performed better in the cognitive task after being treated with the labor market topic and the threshold payment. We do not find that gender plays a role.

[12] A first rough diagnostic is a histogram of the estimated CATE. If the histogram is concentrated at a point, it indicates that the causal forest was not able to detect any heterogeneity. Appendix B Figure B.1 displays the histograms for the main treatment cells. We also use the best linear predictor method, which seeks to fit the CATE as a linear function of the held-out causal forest estimates. It gives a test to assess whether or not the treatment heterogeneity estimates are well calibrated, see (Athey & Wager, 2019). Appendix B Table B.15 summarizes the results.

[13] Appendix B Table B.17 tests the statistical difference of ATE between quartile one and the other quartiles.

Figure 8: ATE within CATE rankings - Cognitive Performance
Labor Market - Threshold



Note: Average treatment effects of the labor market treatment on cognitive performance under the threshold payment scheme, grouped by the quartiles based on the conditional average treatment estimated using the causal forest algorithm. N = 552. (see Table B.16).

We can also pin down other characteristics that seem relevant to generating positive effects from the treatment. First, we find that financial situation and family background play an important role. On average, participants in the highest quartile are less likely to claim financial struggles and rely on their salary. They are more likely to be non-migrants, have highly educated parents, and have both their parents working. They are also less likely to have a family member who lost their job due to the pandemic. Second, we observe that participants who were more socially active during the lockdown are more likely to benefit from the treatment. Students in the highest quartile state not having passed the lockdown alone and claim seeing friends and going to the university more often than those in the lowest quartile. Finally, being able to switch from task to task easily (cognitive flexibility) and having a lower locus of control seem related to performing better.

4.4 Other outcomes: Cognitive reasoning, risk-taking and Maximum WTP

We do not find a significant effect of the topic treatments on cognitive reasoning or risk-taking (see Table A.12 columns 1 to 4). We also verify if the treatment topics affect the willingness to pay for a lottery ticket for the coaching session (Table A.12 columns 5 and 6). We do not find any significant effect (though the coefficients are positive as expected) when controlling only for baseline characteristics. Once we include the extended controls, we find a positive effect (significant only

at 10 percent) of the labor market treatment. Going more into detail, we look at the option that participants chose as their most preferred choice for the coaching session. We find that the labor market treatment significantly increases the likelihood to choose the module "interview simulation" by almost 20%, compared to the control mean. We do not observe any other significant effect for either of the two treatments. Table **??** summarizes the results.

Table 1: Causal Forest: Cognitive Performance - Labor Market and Threshold treatment

Variable	Highest quartile	Lowest quartile	Diff.	P-values
	Panel A. Baseline controls			
Age	20.292	22.768	-2.476	0.00***
Woman	0.728	0.703	0.025	0.65
Scholarship	0.338	0.543	-0.205	0.00***
1st year student	0.412	0.283	0.129	0.02**
Close to labor market	0.066	0.188	-0.122	0.00***
Fatigued	0.794	0.804	-0.010	0.83
First round matrices	1.912	1.862	0.049	0.75
	Panel B. Field of study			
Health Sciences	0.096	0.080	0.016	0.64
Arts and Languages	0.154	0.116	0.038	0.35
Law, Economics, Management	0.228	0.268	-0.040	0.44
Science and Technology	0.272	0.370	-0.098	0.08*
Humanities and Social Science	0.243	0.167	0.076	0.12
	Panel C. Financial Situation			
Having financial struggles	0.110	0.377	-0.267	0.00***
Can afford extra expenses	0.882	0.717	0.165	0.00***
Having own salary	0.103	0.203	-0.100	0.02**
	Panel D. Expectations			
Low prob. success career	0.125	0.196	-0.071	0.11
Low prob. success studies	0.456	0.428	0.028	0.64
Pessimistic about the next 5 years	0.353	0.370	-0.017	0.78
Pressure to have diploma	0.404	0.333	0.071	0.22
Pressure to have good grades	0.199	0.254	-0.055	0.28
	Panel E. Family Background			
Migrant	0.015	0.319	-0.304	0.00***
Living alone	0.279	0.268	0.011	0.83
Father university degree	0.529	0.174	0.355	0.00***
Mother university degree	0.787	0.080	0.707	0.00***
Both parents work	0.787	0.435	0.352	0.00***
	Panel F. Mental Health			
Depression	0.456	0.551	-0.095	0.12
Anxiety	0.412	0.486	-0.074	0.22
	Panel G. Covid-19 Experience			
Had Covid-19	0.103	0.043	0.059	0.06*
Family member had Covid-19	0.279	0.283	-0.003	0.95
Personal traumatic experience	0.206	0.268	-0.062	0.23
Family member lost job	0.154	0.283	-0.128	0.01**
Positive attitude tw vaccination	0.463	0.304	0.159	0.01***
	Panel H. Covid-19 Social experience:			
Lock-down alone	0.022	0.094	-0.072	0.01**
Seeing friends	2.941	0.964	1.977	0.00***
Going to the university	1.493	0.899	0.594	0.00***
	Panel I. Self- perception			
Cognitive flexibility	3.110	3.783	-0.672	0.00***
Cognitive control	2.874	3.036	-0.162	0.16
Locus of control	17.930	19.721	-1.790	0.00***

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. This table compares the characteristics of students that belong to the highest and lowest quartile of the CATE distribution. Those in the highest quartile are predicted to have a positive treatment effect while those in the lowest quartile are predicted to have no treatment effect. See Appendix B.

5 Mechanisms and discussion

We find that highlighting the potential negative consequences of the Covid pandemic on students' job prospects increases their performance on a cognitive functioning test when the financial reward is conditional on reaching a threshold. Conversely, when the economic reward is not linked to an explicit goal, we find that students' performance is not affected. In this section, we explore possible mechanisms underlying these results. We also discuss the potential consequences and external validity of the results.

5.1 Positive effect of goal-based payment and a topic with scope for action

The labor market topic is a reminder of future uncertainty against which students can take potential actions, e.g. through academic effort. The labor market topic might thus generate a willingness to make an extra effort. It can be motivating for those who have the scope to deal with the consequences and improve their changes. This interpretation is underlined by the observation that the results are driven by those not too close to labor market entry. We also observe that those in the labor market treatment state a higher importance of a well-paid job after university (see Table A.13).

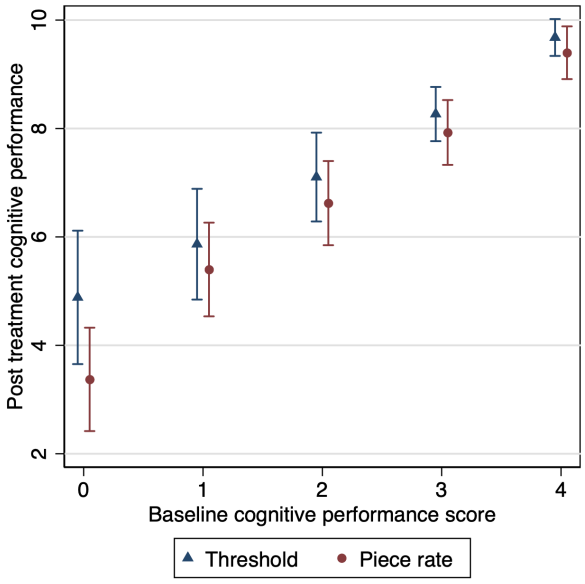
This motivation effect might have been picked up by the more challenging or motivating threshold-payment scheme. Payment schemes that provide explicit and achievable goals appear to enhance motivation and performance more than schemes where payment is linked to the individual unit of output (Bonner et al., 2000). Indeed, the threshold in our experiment is achievable: Overall, 76% of the participants reach the cap of 5 correct answers. Even among those that do not have any correct answer in the training task, 38% reach the threshold. Some psychological studies have also shown that a higher stress level can increase performance by transforming a threat into a challenge and generating a 'challenge state' (Kassam et al., 2009; Arnsten, 2009). Furthermore, those in the threshold condition state a higher importance of good grades and a well-paying career (see Table A.13)[14]. This suggests that the threshold payment scheme puts students in a "performance mindset" – similar to an exam situation.

The threshold payment should be especially motivating for those that are close to reaching the

[14]For the other two outcomes - the importance of finishing their studies and having a job they like we do not find any significant differences. However, nearly everyone rated them at the highest level (indispensable).

threshold. This is what we see after the labor market topic, illustrated in Figure 9. The treatment effect is the strongest among those with a lower "Baseline cognitive performance score". This score measures how many matrices they got correct in the incentivized training task before they saw the treatment.

Figure 9: Treatment effect of the Labor Market Treatment on Cognitive Performance by baseline cognitive ability



Note: Illustrates the predicted cognitive performance as a function of the baseline cognitive performance score, comparing the threshold payment and the piece-rate payment for control. The regression includes pre-registered baseline controls: Gender, age, field of study, year of study, and scholarship recipient. N=1503.

Finally, the labor market topic might have increased the salience of financial resources as some reflection questions were asked about the student's economic situation, enhancing the motivating effect of the threshold payment. Yet, we find that students with a financially-stable background and further away from labor market entry benefit most from the motivational boost. For those financially vulnerable or those whose entry conditions into the labor market are already set the motivational character of the goal-based payment might be counteracted by worries about their financial future.

5.2 No or weakly negative effect of a topic with no scope for action

Contrary to the labor market topic, the mental health topic exposed students to a current and certain 'state of affairs'. We find that the mental health topic not only worsens their depression score (Table A.14 column 1) but also decreases the participant's locus of control (Table A.14 columns 2 - 5) – their belief that their outcomes are mainly driven by their actions rather than chance and circumstances. Indeed, in the piece-rate treatment, we find a weak negative effect among different subgroups, including those with a high depression score. Thus, there is a potential detrimental effect of this deterministic topic on the most vulnerable.

For the threshold payment, we see an average null effect of the mental health topic, and no subgroup with a positive or negative effect. The threshold payment is only expected to be motivating if people believe that they will reach the goal. The mental health topic might have decreased the confidence that they can reach the goal through increased effort. Moreover, if the treatment topic taxes the 'mental bandwidth' and thus increases the cost of effort, people that are not sure to pass the threshold might decrease their effort because the expected benefits are lower. This could counteract the otherwise motivating effect of the threshold payment.

5.3 Potential consequences and external validity

The current literature on financial strain stresses an inequality-deepening or poverty-preserving effect: Those that are financially vulnerable experience a negative effect on cognitive performance when faced with financially worrying tasks or situations (Mani et al. (2013), Duquennois (2022)).

In our case, by combining a troubling topic with scope for action and a goal-based payment scheme, we observe the opposite effect, yet only for those with larger financial resources. Those that are from a financially-stable background seem to be able to draw motivation from future uncertainty. This seems not to be the case for those that are financially vulnerable. During periods when students are faced with extremely worrisome news, we might thus expect those that are already facing additional hurdles due to financial vulnerability to perform worse than financially-stable students. The latter might perform better under these circumstances because they can see the opportunities in the situation, and not worry as much about the negative consequences. This leads to differences in the performance level that are due to pre-existing unequal financial situations,

deepening or preserving pre-existing inequalities.

The question arises if these results hold beyond a sample of university students. Indeed, the goal-based payment scheme seems to put students in an "exam mindset" which might not be the same for older adults. Also, we can confront students with a future worrying topic that they can still individually affect through effort. Those that are already in the labor market might not see the opportunities when reflecting on an uncertain labor market. Yet, it is conceivable that a topic such as climate change which for many is still a topic with consequences in the future that one can prepare for, can be seen as a challenge by some - who then get motivated to become active - and as a distracting worry by others.

Finally, we do not find a negative average effect of our treatments, even the mental health one. We only find a weak effect among those that are mentally vulnerable. However, when students were invited to sign up and when they started the survey, they were told that it included questions about the pandemic. Thus, those that were the most vulnerable might have not taken part in the survey. Furthermore, we selected topics that were negative but correct and not sensationalist. On social media, people might be confronted with much worse framing of these Covid-related topics. We also just test the effect of one such reminder of a topic that they might have already heard a lot about and contemplated on several occasions. Therefore, we most likely find a lower bound of the effect.

6 Conclusion

We study the effect of highlighting negative Covid-19-related topics on cognitive performance under different payment schemes. Ongoing reminders of negative events that students are faced with during this pandemic can increase their worries. We do not find evidence that one such reminder has a negative effect on cognitive performance, nor on cognitive reasoning or risk aversion. This is only partly reassuring since our priming might be relatively mild compared to the extent to which students' lives have been impacted by the pandemic. Yet it might indicate that the majority can still concentrate and "perform" when it counts.

Interestingly, we find a positive effect of highlighting the negative labor market consequences when participants face a goal-based payment structure. The payment structure combined with the

topic stressing labor market entry and financial resources seems to be motivating particularly for those who are from a better-off background.

Our results stress that different types of worries might affect cognitive performance differently. For some, they might motivate them to do better, while they can depress others. Also, the incentive structure that people face can interact with these worries. We find that worries that concern the future and where there is still scope for action can improve cognitive performance among those with larger financial resources. Pandemic-related worries, as well as other worries, might therefore exacerbate pre-existing inequalities.

References

Achtziger, A., Alós-Ferrer, C., & Ritschel, A. (2020). Cognitive load in economic decisions. *University of Zurich, Department of Economics, Working Paper*(354).

Ariely, D., Gneezy, U., Loewenstein, G., & Mazar, N. (2009). Large stakes and big mistakes. *The Review of Economic Studies*, *76*(2), 451–469.

Arnsten, A. F. (2009). Stress signalling pathways that impair prefrontal cortex structure and function. *Nature Reviews Neuroscience*, *10*(6), 410–422.

Athey, S., & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, *113*(27), 7353–7360.

Athey, S., & Wager, S. (2019). Estimating treatment effects with causal forests: An application. *Observational Studies*, *5*(2), 37–51.

Baumeister, R. F. (1984). Choking under pressure: self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology*, *46*(3), 610.

Belle, N., & Cantarelli, P. (2015). Monetary incentives, motivation, and job effort in the public sector: An experimental study with italian government executives. *Review of Public Personnel Administration*, *35*(2), 99–123.

Bogliacino, F., Codagnone, C., Montealegre, F., Folkvord, F., Gómez, C., Charris, R., ... Veltri, G. A. (2021). Negative shocks predict change in cognitive function and preferences: Assessing the negative affect and stress hypothesis. *Scientific Reports*, *11*(1), 1–10.

Bonner, S. E., Hastie, R., Sprinkle, G. B., & Young, S. M. (2000). A review of the effects of financial incentives on performance in laboratory tasks: Implications for management accounting. *Journal of Management Accounting Research*, *12*(1), 19–64.

Bonner, S. E., & Sprinkle, G. B. (2002). The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. *Accounting, Organizations and Society*, *27*(4-5), 303–345.

Camerer, C. F., & 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), 7–42.

Canavari, M., Drichoutis, A. C., Lusk, J. L., & Nayga Jr, R. M. (2019). How to run an experimental auction: A review of recent advances. *European Review of Agricultural Economics*, *46*(5), 862–922.

Carvalho, L. S., Meier, S., & Wang, S. W. (2016). Poverty and economic decision-making: Evidence from changes in financial resources at payday. *American Economic Review*, *106*(2), 260–84.

Castro, J. F., Yamada, G., Contreras, H., Linares, F., & Watson, H. (2021). Tracking the pressure: The effect of monetary incentives on effort, emotions, and performance. *Journal of Neuroscience, Psychology, and Economics*, *14*(4), 234.

Chierchia, G., Fuhrmann, D., Knoll, L. J., Pi-Sunyer, B. P., Sakhardande, A. L., & Blakemore, S.-J. (2019). The matrix reasoning item bank (mars-ib): novel, open-access abstract reasoning items for adolescents and adults. *Royal Society open science*, *6*(10), 190232.

Cole, J. S., Bergin, D. A., & Summers, J. (2018). A lottery improves performance on a low-stakes test for males but not females. *Assessment in Education: Principles, Policy & Practice*, *25*(5), 488–503.

Dalton, P. S., Nhung, N., & Rüschenpöhler, J. (2020). Worries of the poor: The impact of financial burden on the risk attitudes of micro-entrepreneurs. *Journal of Economic Psychology*, *79*, 102198.

Deck, C., & Jahedi, S. (2015). The effect of cognitive load on economic decision making: A survey and new experiments. *European Economic Review*, *78*, 97–119.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168.

Duquennois, C. (2022). Fictional money, real costs: Impacts of financial salience on disadvantaged students. *American Economic Review*, *112*(3), 798–826.

Duttweiler, P. C. (1984). The internal control index: A newly developed measure of locus of control. *Educational and Psychological Measurement*, *44*(2), 209–221.

Earley, P. C., & Lituchy, T. R. (1991). Delineating goal and efficacy effects: A test of three models. *Journal of Applied Psychology*, *76*(1), 81.

Etheridge, B., & Spantig, L. (2022). The gender gap in mental well-being at the onset of the Covid-19 pandemic: Evidence from the UK. *European Economic Review*, *145*, 104114.

Farbmacher, H., Kögel, H., & Spindler, M. (2021). Heterogeneous effects of poverty on attention. *Labour Economics*, *71*, 102028.

Fetzer, T., Hensel, L., Hermle, J., & Roth, C. (2020). Coronavirus perceptions and economic anxiety. *Review of Economics and Statistics*, 1–36.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42.

Gignac, G. E. (2018). A moderate financial incentive can increase effort, but not intelligence test performance in adult volunteers. *British Journal of Psychology*, *109*(3), 500–516.

Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, *112*(2), 631–645.

González-Arango, F., Corredor, J., López-Ardila, M. A., Contreras-González, M. C., Herrera-Santofimio, J., & González, J. J. (2021). The duality of poverty: a replication of mani et al.(2013) in colombia. *Theory and Decision*, 1–35.

Haushofer, J., & Fehr, E. (2014). On the psychology of poverty. *science*, *344*(6186), 862–867.

Hickman, D. C., & Metz, N. E. (2015). The impact of pressure on performance: Evidence from the pga tour. *Journal of Economic Behavior & Organization*, *116*, 319–330.

Hinz, A., Daig, I., Petrowski, K., & Brähler, E. (2012). Mood in the german population: norms of the multidimensional mood questionnaire mdbf. *Psychotherapie, Psychosomatik, medizinische Psychologie*, *62*(2), 52–57.

Jenkins Jr, G. D., Mitra, A., Gupta, N., & Shaw, J. D. (1998). Are financial incentives related to performance? a meta-analytic review of empirical research. *Journal of Applied Psychology*, *83*(5), 777.

Kassam, K. S., Koslov, K., & Mendes, W. B. (2009). Decisions under distress: Stress profiles influence anchoring and adjustment. *Psychological Science*, *20*(11), 1394–1399.

Kaur, S., Mullainathan, S., Oh, S., & Schilbach, F. (2021). *Do financial concerns make workers less productive?* (Tech. Rep.). National Bureau of Economic Research.

Kroenke, K., & Spitzer, R. L. (2002). *The phq-9: a new depression diagnostic and severity measure* (Vol. 32) (No. 9). SLACK Incorporated Thorofare, NJ.

Levitt, S. D., List, J. A., & Sadoff, S. (2016). *The effect of performance-based incentives on educational achievement: Evidence from a randomized experiment* (Tech. Rep.). National Bureau of Economic Research.

Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. *Science*, *341*(6149), 976–980.

Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2020). Scarcity and cognitive function around payday: A conceptual and empirical analysis. *Journal of the Association for Consumer Research*, *5*(4), 365–376.

McGinty, E. E., Presskreischer, R., Han, H., & Barry, C. L. (2020). Psychological distress and loneliness reported by us adults in 2018 and april 2020. *Jama*, *324*(1), 93–94.

McLaughlin, B., Gotlieb, M. R., & Mills, D. J. (2022). Caught in a dangerous world: Problematic news consumption and its relationship to mental and physical ill-being. *Health Communication*, *0*(0), 1-11. (PMID: 35999665) doi: 10.1080/10410236.2022.2106086

Mullainathan, S., & Shafir, E. (2013). Decision making and policy in contexts of poverty. In *The behavioral foundations of public policy* (pp. 281–298). Princeton University Press.

OECD. (2020). *Youth and covid-19: Response, recovery and resilience* (Tech. Rep.).

Pierce, M., Hope, H., Ford, T., Hatch, S., Hotopf, M., John, A., ... others (2020). Mental health before and during the covid-19 pandemic: a longitudinal probability sample survey of the uk population. *The Lancet Psychiatry*, *7*(10), 883–892.

Raven, J., et al. (2003). Raven progressive matrices. In *Handbook of nonverbal assessment* (pp. 223–237). Springer.

Romano, J. P., & Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, *73*(4), 1237–1282.

Santé publique France. (2021, December). *Coviprev: une enquête pour suivre l'évolution des comportements et de la santé mentale pendant l'épidémie* (Tech. Rep.). French Health Ministry. Retrieved from https://www.santepubliquefrance.fr/etudes-et-enquetes/coviprev-une-enquete-pour-suivre-l-evolution-des-comportements-et-de-la-sante-mentale-pendant-l-epidemie-de-covid-19

Schilbach, F., Schofield, H., & Mullainathan, S. (2016). The psychological lives of the poor. *American Economic Review*, *106*(5), 435–40.

Shah, A. K., Mullainathan, S., & Shafir, E. (2012). Some consequences of having too little. *Science*, *338*(6107), 682–685.

Thompson, R. R., Jones, N. M., Holman, E. A., & Silver, R. C. (2019). Media exposure to mass violence events can fuel a cycle of distress. *Science advances*, *5*(4), eaav3502.

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113*(523), 1228–1242.

Wright, P. M. (1990). Monetary incentives and task experience as determinants of spontaneous goal setting, strategy development, and performance. *Human Performance*, *3*(4), 237–258.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, *67*(6), 361–370.

A Appendix Additional Tables

Table A.1: Balance table on baseline characteristics

		Labor Market vs Controls		Mental Health vs Controls		Threshold vs Piece-rate	
Variable	Mean	Diff	P-values	Diff	P-values	Diff	P-values
Woman	0.66	0.05	0.08*	0.01	0.65	0.01	0.60
Scholarship	0.45	-0.00	0.91	0.05	0.14	-0.00	0.86
Age	21.57	0.04	0.78	-0.10	0.54	-0.19	0.14
First Round Matrices	2.04	0.03	0.72	-0.08	0.29	-0.08	0.22
- Level of study:							
Undergrad	0.70	-0.00	0.89	-0.00	0.97	0.02	0.37
Master and Engineers	0.27	0.00	0.93	-0.01	0.77	-0.02	0.44
PhD	0.03	0.00	0.87	0.01	0.36	-0.00	0.68
- Field of study:							
Arts and Languages	0.14	-0.00	0.85	0.00	0.84	-0.01	0.68
Health Sciences	0.11	-0.00	0.86	-0.00	0.89	-0.01	0.55
Law, Economics, Management	0.25	0.01	0.66	0.01	0.72	0.01	0.82
Humanities and Social Sciences	0.17	-0.02	0.34	0.01	0.83	0.01	0.72
Science and Technology	0.34	0.02	0.55	-0.02	0.58	0.00	0.85
Observations	1503						
Joint orthogonality test			0.88		0.82		0.89

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. P-values reported are from a t-test of equality of means.

Table A.2: Balance table on baseline characteristics under piece-rate payment

		Labor Market vs Controls		Mental Health vs Controls	
Variable	Mean	Diff	P-values	Diff	P-values
Woman	0.66	0.06	0.13	0.01	0.79
Scholarship	0.45	-0.01	0.81	0.05	0.25
Age	21.66	-0.07	0.74	-0.19	0.38
First Round Matrices	2.08	-0.00	0.98	-0.21	0.05**
- Level of study:					
Undergrad	0.69	0.06	0.18	0.04	0.34
Master and Engineers	0.28	-0.06	0.14	-0.03	0.42
PhD	0.03	0.00	0.77	-0.01	0.62
- Field of study:					
Arts and Languages	0.14	0.01	0.64	0.03	0.29
Health Sciences	0.11	-0.01	0.78	0.00	0.98
Law, Economics, Management	0.25	0.02	0.69	-0.00	0.97
Humanities and Social Sciences	0.16	-0.02	0.48	-0.02	0.52
Science and Technology	0.34	0.00	0.97	-0.01	0.79
Observations	779				
Joint orthogonality test			0.82		0.73

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. P-values reported are from a t-test of equality of means.

Table A.3: Balance table on baseline characteristics under threshold payment

		Labor Market vs Controls		Mental Health vs Controls	
Variable	Mean	Diff	P-values	Diff	P-values
Woman	0.67	0.04	0.36	0.02	0.70
Scholarship	0.44	0.00	0.94	0.04	0.36
Age	21.47	0.17	0.44	0.01	0.97
First Round Matrices	2.00	0.06	0.58	0.05	0.65
- *Level of study:*					
Undergrad	0.72	-0.07	0.10*	-0.05	0.26
Master and Engineers	0.26	0.07	0.09*	0.02	0.64
PhD	0.02	-0.00	0.93	0.03	0.07*
- *Level of study:*					
Arts and Languages	0.13	-0.02	0.45	-0.03	0.41
Health Sciences	0.10	0.00	0.97	-0.01	0.82
Law, Economics, Management	0.25	0.01	0.82	0.02	0.58
Humanities and Social Sciences	0.17	-0.02	0.52	0.03	0.35
Science and Technology	0.34	0.04	0.42	-0.02	0.60
Observations	724				
Joint orthogonality test			0.83		0.54

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. P-values reported are from a t-test of equality of means.

Table A.4: Effect of the treatment on emotional states.

	Dependent variable:					
	Feeling good		Feeling awake		Feeling calm	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Animal Welfare Control						
Before article	-0.216	-0.316	0.296	0.125	0.217	0.0946
	(0.345)	(0.321)	(0.316)	(0.232)	(0.347)	(0.313)
Labor Market	-1.041*	-1.075**	-0.531	-0.595*	-0.837*	-0.871*
	(0.418)	(0.396)	(0.379)	(0.292)	(0.409)	(0.381)
Mental Health	-1.419***	-1.479***	-0.707	-0.793**	-1.411***	-1.512***
	(0.411)	(0.375)	(0.393)	(0.296)	(0.413)	(0.387)
Observations	1085	1085	1094	1094	1090	1090
Control Mean	16.00	16.00	12.06	12.06	13.00	13.00
Panel B: Space Program Control						
Before article	-0.338	-0.442	0.627*	0.404	0.0896	-0.0156
	(0.324)	(0.280)	(0.312)	(0.219)	(0.340)	(0.296)
Labor Market	-1.378***	-1.297***	-0.358	-0.301	-1.010*	-0.920*
	(0.403)	(0.367)	(0.377)	(0.286)	(0.405)	(0.369)
Mental Health	-1.755***	-1.664***	-0.518	-0.442	-1.548***	-1.461***
	(0.397)	(0.345)	(0.391)	(0.282)	(0.407)	(0.367)
Observations	1101	1101	1113	1113	1109	1109
Control Mean	16.37	16.37	12.12	12.12	12.98	12.98
Panel C: Both Controls						
Before Article	-0.288	-0.376	0.439	0.240	0.122	0.0193
	(0.250)	(0.223)	(0.234)	(0.167)	(0.257)	(0.225)
Labor Market	-1.230***	-1.215***	-0.450	-0.463	-0.936**	-0.916**
	(0.355)	(0.331)	(0.325)	(0.250)	(0.348)	(0.321)
Mental Health	-1.624***	-1.597***	-0.619	-0.625*	-1.500***	-1.505***
	(0.349)	(0.307)	(0.341)	(0.251)	(0.351)	(0.324)
Observations	1482	1482	1497	1497	1492	1492
Control Mean	16.17	16.17	12.09	12.09	12.99	12.99
Baseline Controls	yes	yes	yes	yes	yes	yes
Extended controls	no	yes	no	yes	no	yes

Note: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Scores are based on four questions for each mood (two positively phrased, two negatively). Minimum possible 4, maximum possible 24. Baseline controls include gender, age, field of study, year of study, number of correct matrices in the first round, and scholarship recipient. Extended controls include the day of the survey, week of the survey, time of the survey, level of fatigue, whether they are a French native speaker, whether they were born abroad, whether both their parents were born abroad, color blindness.

Table A.5: Effect of the treatment on cognitive performance.

Dependent variable: Cognitive performance				
	Piece-rate payment		Threshold payment	
	(1)	(2)	(3)	(4)
Labor Market	-0.239	-0.221	0.460**	0.425*
	(0.203)	(0.206)	(0.218)	(0.223)
Mental Health	-0.227	-0.188	-0.0403	-0.0967
	(0.206)	(0.204)	(0.222)	(0.227)
Observations	779	779	724	724
Control Mean	6.97	6.97	6.64	6.64
Baseline Controls	yes	yes	yes	yes
Extended controls	no	yes	no	yes

Note: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the number of correct matrices. Baseline controls include gender, age, field of study, year of study, number of correct matrices in the first round, and scholarship recipient. Extended controls include the day of the survey, week of the survey, time of the survey, level of fatigue, whether they are a French native speaker, whether they were born abroad, whether both their parents were born abroad, color blindness.

Table A.6: Effect of the treatment on cognitive performance with interaction terms.

Dependent variable: Cognitive performance		
	(1)	(2)
Labor Market	-0.246	-0.226
	(0.202)	(0.204)
Mental Health	-0.228	-0.206
	(0.205)	(0.203)
Threshold	-0.118	-0.0917
	(0.160)	(0.162)
Labor Market X Threshold	0.696**	0.617**
	(0.297)	(0.298)
Mental Health X Threshold	0.190	0.137
	(0.301)	(0.301)
Observations	1503	1503
Control Mean	6.97	6.97
Baseline Controls	yes	yes
Extended controls	no	yes

Note: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the number of correct matrices. Baseline controls include gender, age, field of study, year of study, number of correct matrices in the first round, and scholarship recipient. Extended controls include the day of the survey, week of the survey, time of the survey, level of fatigue, whether they are a French native speaker, whether they were born abroad, whether both their parents were born abroad, color blindness.

Table A.7: Pre-registered heterogeneity: Treatment effect on cognitive performance under piece-rate payment.

	Gender		Scholarship		Level of study		Labor Market	
	Woman	Man	With	Without	1st year	Not 1st year	Close	Not close
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Labor Market	-0.170	-0.339	-0.219	-0.205	-0.014	-0.360	0.246	-0.240
	(0.474)	(0.386)	(0.466)	(0.457)	(0.968)	(0.155)	(0.666)	(0.268)
	[0.872]	[0.872]	[0.803]	[0.803]	[0.969]	[0.473]	[0.902]	[0.687]
Mental Health	-0.195	-0.239	0.243	-0.634	-0.150	-0.254	-0.255	-0.123
	(0.447)	(0.493)	(0.405)	(0.028)**	(0.672)	(0.315)	(0.662)	(0.582)
	[0.872]	[0.872]	[0.803]	[0.096]*	[0.897]	[0.685]	[0.902]	[0.902]
Observations	513	266	349	430	257	522	114	665
Control Mean	6.83	6.71	6.50	7.03	6.54	6.91	7.20	6.72
	Field		Depression		Anxiety		Emotion qnt	
	Health Sciences	Others	Above median	Below median	Above median	Below median	After	Before
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Labor Market	-0.888	-0.165	-0.330	-0.074	-0.339	-0.190	-0.418	-0.070
	(0.057)*	(0.451)	(0.237)	(0.801)	(0.264)	(0.479)	(0.188)	(0.785)
	[0.165]	[0.682]	[0.520]	[0.940]	[0.590]	[0.713]	[0.456]	[0.944]
Mental Health	-1.252	-0.106	-0.599	0.096	-0.497	-0.069	-0.463	0.020
	(0.017)**	(0.635)	(0.038)**	(0.749)	(0.149)	(0.791)	(0.136)	(0.941)
	[0.083]*	[0.682]	[0.136]	[0.940]	[0.455]	[0.782]	[0.432]	[0.944]
Observations	87	692	379	400	309	470	372	407
Control Mean	7.47	6.71	6.72	6.86	6.47	7.00	6.75	6.83

Note: P-values in parentheses. Robust standard errors. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Romano & Wolf (2005) adjusted P-values for multiple hypothesis testing in square brackets with 1,000 replications. The dependent variable is the number of correct matrices. All specifications include baseline controls: gender, age, field of study, year of study, number of correct matrices in the first round, and scholarship recipient.

Table A.8: Pre-registered heterogeneity: Treatment effect on cognitive performance under threshold payment.

	Gender		Scholarship		Level of study		Labor Market	
	Woman	Man	With	Without	1st year	Not 1st year	Close	Not close
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Labor Market	0.602	0.192	0.459	0.488	0.364	0.539	0.189	0.521
	(0.021)**	(0.637)	(0.172)	(0.095)*	(0.341)	(0.042)**	(0.715)	(0.030)**
	[0.094]*	[0.943]	[0.440]	[0.334]	[0.705]	[0.153]	[0.973]	[0.135]
Mental Health	-0.041	-0.036	-0.234	0.114	-0.153	0.011	-0.074	-0.032
	(0.879)	(0.930)	(0.495)	(0.697)	(0.666)	(0.968)	(0.908)	(0.894)
	[0.981]	[0.981]	[0.771]	[0.771]	[0.877]	[0.972]	[0.985]	[0.985]
Observations	486	238	321	403	251	473	96	628
Control Mean	6.92	6.47	6.41	7.07	6.57	6.88	7.21	6.71
	Field		Depression		Anxiety		Emotion qnt	
	Health Sciences	Others	Above median	Below median	Above median	Below median	After	Before
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Labor Market	1.411	0.375	0.498	0.476	0.564	0.372	0.362	0.534
	(0.063)**	(0.099)*	(0.145)	(0.099)*	(0.086)*	(0.187)	(0.248)	(0.081)*
	[0.242]	[0.282]	[0.352]	[0.342]	[0.304]	[0.452]	[0.569]	[0.274]
Mental Health	0.165	-0.060	-0.368	0.303	-0.373	0.194	-0.138	0.115
	(0.839)	(0.795)	(0.271)	(0.314)	(0.318)	(0.482)	(0.660)	(0.721)
	[0.945]	[0.945]	[0.436]	[0.436]	[0.508]	[0.508]	[0.867]	[0.867]
Observations	74	650	345	379	307	417	365	359
Control Mean	6.99	6.75	6.89	6.67	6.89	6.70	6.88	6.67

Note: P-values in parentheses. Robust standard errors. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Romano & Wolf (2005) adjusted P-values for multiple hypothesis testing in square brackets with 1,000 replications. The dependent variable is the number of correct matrices. All specifications include baseline controls: gender, age, field of study, year of study, number of correct matrices in the first round, and scholarship recipient.

Table A.9: Pre-registered heterogeneity: Treatment effect on cognitive performance with interaction terms.

	Gender		Scholarship		Level of study		Labor Market	
	Woman	Man	With	Without	1st year	Not 1st year	Close	Not close
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Labor Market	-0.204	-0.343	-0.255	-0.225	0.016	-0.376	-0.043	-0.256
	(0.388)	(0.375)	(0.395)	(0.412)	(0.962)	(0.135)	(0.942)	(0.234)
Mental Health	-0.193	-0.266	0.208	-0.628**	-0.130	-0.252	-0.571	-0.140
	(0.451)	(0.443)	(0.476)	(0.030)	(0.715)	(0.319)	(0.315)	(0.532)
Threshold	-0.108	-0.051	-0.047	-0.182	0.242	-0.303	-0.164	-0.098
	(0.584)	(0.855)	(0.849)	(0.386)	(0.395)	(0.122)	(0.743)	(0.563)
Labor Market X Threshold	0.771**	0.501	0.695	0.694*	0.319	0.923**	0.232	0.754**
	(0.028)	(0.372)	(0.123)	(0.080)	(0.538)	(0.011)	(0.767)	(0.020)
Mental Health X Threshold	0.107	0.248	-0.423	0.756*	-0.004	0.262	0.493	0.106
	(0.773)	(0.642)	(0.345)	(0.064)	(0.994)	(0.490)	(0.560)	(0.744)
Observations	999	504	670	833	508	995	210	1293
Control Mean	6.83	6.71	6.50	7.03	6.54	6.91	7.20	6.72
	Field		Depression		Anxiety		Emotion qnt	
	Health Sciences	Others	Above median	Below median	Above median	Below median	After	Before
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Labor Market	-0.792*	-0.171	-0.353	-0.125	-0.357	-0.209	-0.430	-0.102
	(0.086)	(0.432)	(0.205)	(0.670)	(0.246)	(0.442)	(0.169)	(0.690)
Mental Health	-1.197**	-0.108	-0.619**	0.115	-0.547	-0.029	-0.477	0.003
	(0.017)	(0.628)	(0.031)	(0.700)	(0.108)	(0.911)	(0.124)	(0.992)
Threshold	-1.358***	0.023	-0.098	-0.164	0.125	-0.296	-0.071	-0.174
	(0.006)	(0.889)	(0.653)	(0.483)	(0.614)	(0.158)	(0.763)	(0.433)
Labor Market X Threshold	2.166**	0.534*	0.822*	0.611	0.926**	0.566	0.794*	0.629
	(0.012)	(0.090)	(0.064)	(0.135)	(0.041)	(0.144)	(0.073)	(0.115)
Mental Health X Threshold	1.382	0.051	0.230	0.209	0.179	0.206	0.320	0.118
	(0.131)	(0.874)	(0.597)	(0.621)	(0.720)	(0.588)	(0.462)	(0.780)
Observations	161	1342	724	779	616	887	737	766
Control Mean	7.47	6.71	6.72	6.86	6.47	7.00	6.75	6.83

Note: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the number of correct matrices. Baseline controls include gender, age, field of study, year of study, number of correct matrices in the first round, and scholarship recipient.

Table A.10: Treatment effect on depression score - Quantile regression.

Dependent variable: Depression score			
	(1)	(2)	(3)
	All	Piece-rate	Threshold
Q25			
Labor Market	0.000	0.794**	-0.485
	(0.432)	(0.396)	(0.507)
Mental Health	0.300	0.688	-0.138
	(0.395)	(0.473)	(0.593)
Q50			
Labor Market	-0.245	0.635	-0.419
	(0.492)	(0.981)	(0.661)
Mental Health	0.532	0.476	0.356
	(0.512)	(0.685)	(1.054)
Q75			
Labor Market	0.655	1.370**	-0.780
	(0.645)	(0.633)	(1.011)
Mental Health	1.794***	2.387***	-0.092
	(0.647)	(0.779)	(1.193)
Observations	1489	767	722
Control Mean	10.74	10.38	11.12

Note: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All specifications include baseline controls: gender, age, field of study, year of study, number of correct matrices in the first round, and scholarship recipient.

Table A.11: Treatment effect with corrected depression score.

Dependent variable: Cognitive performance				
	Piece-rate payment		Threshold payment	
	Above median	Below median	Above median	Below median
	(1)	(2)	(3)	(4)
Labor Market	-0.354	-0.0557	0.501	0.464
	(0.276)	(0.299)	(0.323)	(0.302)
Mental Health	-0.719**	0.216	-0.367	0.305
	(0.295)	(0.293)	(0.333)	(0.301)
Observations	398	381	352	372
Control Mean	7.10	6.83	6.90	6.39

Note: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All specifications include baseline controls: gender, age, field of study, year of study, number of correct matrices in the first round, and scholarship recipient. To account for the negative treatment effect on the Mental Health score, we regressed the treatment topic, payment treatment, and their interaction with the mental health score. Next, we retrieved the residuals, which will contain everything that is not explained by the treatment effects. Afterwards, we split the sample into two groups (below and above the median) using the residuals as the new mental health score.

Table A.12: Treatment effect on cognitive reasoning, risk-taking and maximum WTP.

	Dependent variable:					
	Cognitive reasoning		Risk-taking		Maximum wtp	
	(1)	(2)	(3)	(4)	(5)	(6)
Labor Market	-0.00168	0.00526	-0.0174	-0.0114	0.0874	0.104*
	(0.0633)	(0.0639)	(0.0636)	(0.0637)	(0.0624)	(0.0625)
Mental Health	-0.00463	0.0191	-0.0530	-0.0666	0.0736	0.0774
	(0.0616)	(0.0623)	(0.0628)	(0.0637)	(0.0623)	(0.0633)
Observations	1503	1503	1503	1503	1503	1503
Control Mean	1.20	1.20	1.56	1.56	0.61	0.61
Baseline Controls	yes	yes	yes	yes	yes	yes
Extended controls	no	yes	no	yes	no	yes

Note: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. From columns (1) and (2) the dependent variable, cognitive reasoning, is the number of correct answers to three questions. From columns (3) and (4) the dependent variable, risk-taking, is the amount of money (up to 3€) they invest in a lottery. From columns (5) and (6) the dependent variable, Max. WTP, is the maximum willingness to pay to participate in a lottery for a coaching program. Baseline controls include gender, age, field of study, year of study, number of correct matrices in the first round, and scholarship recipient. Extended controls include the day of the survey, week of the survey, time of the survey, level of fatigue, whether they are a French native speaker, whether they were born abroad, whether both their parents were born abroad, color blindness.

Table A.13: Treatment effect on career goals

	Dependent variable: Importance of having			
	Good grades	Good career	University diploma	Enjoyable job
	(1)	(2)	(3)	(4)
Labor Market	0.0615	0.185**	-0.0214	-0.0416
	(0.101)	(0.0911)	(0.0697)	(0.0602)
Mental Health	0.132	0.0565	-0.0865	-0.0358
	(0.0962)	(0.0919)	(0.0711)	(0.0589)
Threshold	0.176**	0.155**	-0.0151	0.0406
	(0.0794)	(0.0735)	(0.0566)	(0.0448)
Labor Market X Threshold	0.00269	-0.240*	0.0168	0.0824
	(0.145)	(0.127)	(0.100)	(0.0790)
Mental Health X Threshold	-0.131	-0.148	0.153	0.0576
	(0.141)	(0.134)	(0.101)	(0.0795)
Observations	1495	1494	1499	1494
Control Mean	3.06	3.68	4.42	4.64

Note: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Dependent variables range from 1 to 5, where 5 indicates that students rate the respective claim as indispensable. All specifications include baseline controls: gender, age, field of study, year of study, number of correct matrices in the first round, and scholarship recipient.

Table A.14: Treatment effect on depression score and locus of control

	Dependent variable:				
	Depression score	Locus of control	Hard work	Chances are determined	What has to happen will happen
	(1)	(2)	(3)	(4)	(5)
Labor Market	0.752	-0.405	-0.0641	0.0317	0.0519
	(0.474)	(0.274)	(0.0435)	(0.0427)	(0.0416)
Mental Health	1.335***	-0.558**	-0.0802*	0.140***	0.0680*
	(0.476)	(0.260)	(0.0434)	(0.0394)	(0.0393)
Threshold	0.622	-0.134	-0.0203	0.0359	-0.00174
	(0.381)	(0.208)	(0.0343)	(0.0341)	(0.0339)
Labor Market X Threshold	-1.309*	0.428	-0.00746	-0.0220	-0.0505
	(0.702)	(0.386)	(0.0628)	(0.0609)	(0.0607)
Mental Health X Threshold	-1.087	0.253	0.0411	-0.144**	0.00458
	(0.720)	(0.379)	(0.0625)	(0.0588)	(0.0577)
Observations	1489	1453	1503	1503	1503
Control Mean	10.38	19.22	0.64	0.63	0.63

Note: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. This table shows the treatment effect on Depression score and Locus of control. The dependent variable in column (1) is a depression score measured through a short version of the PHG-9. The higher the score, the greater the depression symptoms. Column (2) is the locus of control index (ICI), a higher locus of control indicates that subjects believe they have control over life events. From columns (2) to (4) the dependent variable is a dummy equal to 1 for students who agree with the following statements. Column (3): *To succeed, you have to work hard; success has nothing to do with luck.* Column (4): *The opportunities a person has in life depend on the social conditions in which they live.* Column (5): *I often tell myself that what has to happen will happen somehow.* All specifications include baseline controls: gender, age, field of study, year of study, number of correct matrices in the first round, and scholarship recipient.

B Appendix Causal Forest

Causal Forest is a non-parametric method that allows uncovering heterogeneity in treatment effects (Wager & Athey, 2018). A causal forest is the average of many different casual trees i.e., a data-driven approach to split the data into subgroups that differ in the magnitude of treatment effects (Athey & Imbens, 2016). Importantly, the goal of this method is to obtain accurate estimates for the Conditional Average Treatment Effect (CATE):

$$\tau(x) = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = x] \tag{1}$$

where Y is the outcome of interest, and X is a vector of observable characteristics. A central feature of this method relies on sample splitting, which allows for obtaining valid asymptotic confidence intervals of the parameters estimated. The idea is to use different sub-samples for split selection and estimation. This type of sample splitting is called honesty (Athey & Imbens, 2016).

Following Athey & Wager (2019), we implement the causal forest algorithm as follows:

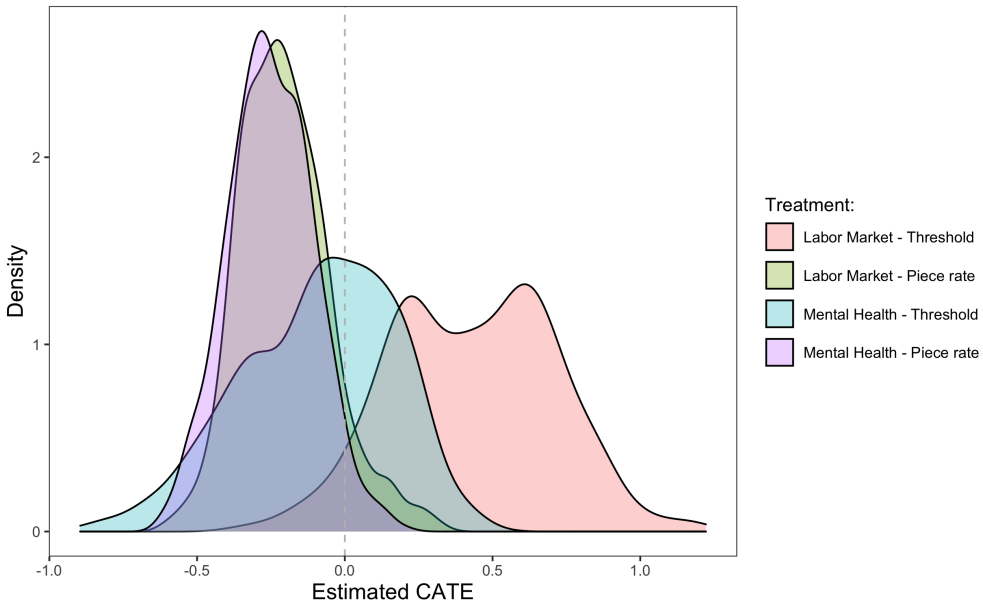
1. We split the data into 5-folds.
2. We take one fold as a hold-out or test data, and we use the remaining folds as training data to fit a CATE model with 20,000 trees.
3. We use the held-out fold to rank the observations into quartiles according to their CATE prediction.
4. We cycle through all the folds by applying the same procedure.[15]
5. We fit 100 causal forests and we average out their prediction to guarantee stability in our results.
6. Next, we concatenate the independent rankings together to study the differences between each rank-group.

We apply the causal forest to predict heterogeneity in treatment effects on the cognitive performance outcome for all possible combinations of treatment cells. We assess the quality of the predicted heterogeneity for our primary treatment cells. The first rough test is a histogram of the

[15]This sample splitting will ensure that the honesty criterion is met so we can obtain valid tests for each quartile.

estimated CATE. Figure B.1 presents the results. Under the piece-rate treatment and regardless of the topic treatment, the distribution of the CATE is very concentrated at a point. This suggests that the variables capture little heterogeneity in the underlying results. Little variation in the CATE means that the causal forest could be underpowered and is unable to detect differences in treatment effects (Athey & Wager, 2019). Contrary, we observe that the histogram for any of the topic treatments under the threshold payment is spread out indicating plausible heterogeneity.

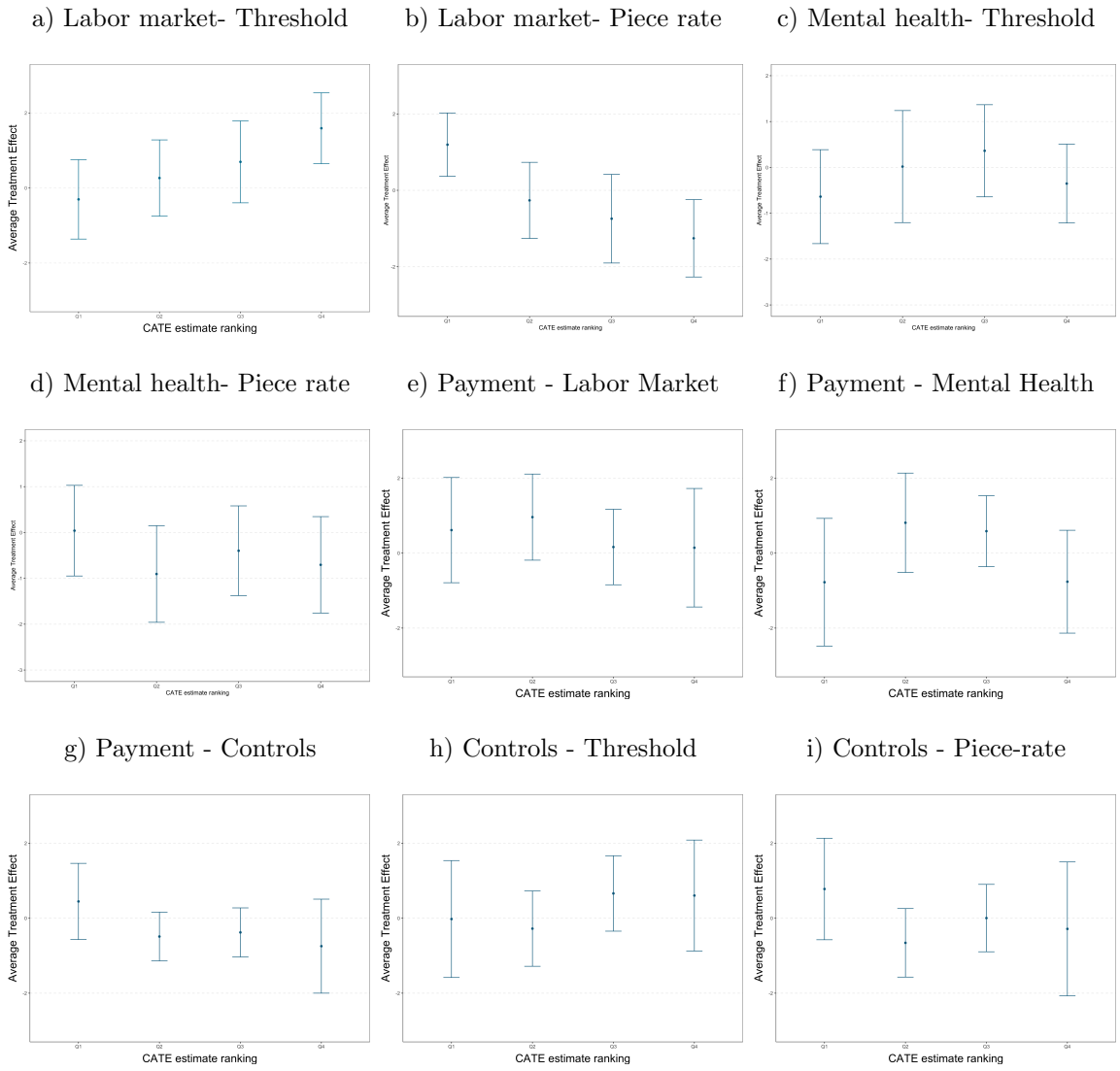
Figure B.1: Histogram Estimated CATE



Distribution of the Conditional Treatment Effects estimated by the causal forest from Wager and Athey (2018). Comparison is the control treatments under the same payment scheme.

Second, Figure B.2 displays the average treatment effect within each predicted rank-group as defined above. Except for the Labor Market-Threshold treatment, the plots are not monotonic or decreasing along the predicted subgroups – meaning that the CATE rankings capture noise rather than heterogeneity.

Figure B.2: Average Treatment Effect on cognitive performance



Note: Average treatment effects of the different treatment cells, grouped by the quartiles based on the conditional average treatment estimated using the causal forest algorithm.

Third, we use the best linear predictor method to assess whether the suggestive heterogeneity found is meaningful and not pure noise. This calibration test seeks to fit the CATE as a linear function of the held-out causal forest estimates. A coefficient of 1 for the "Mean forest prediction" suggest that the prediction produced by the forest is correct, and a coefficient of 1 for "Differential forest prediction" suggest that the forest is capturing heterogeneity (see Athey & Wager (2019) for details on the test). Table B.15 summarizes the results. We find that the only treatment with significant underlying heterogeneity is the labor market treatment under the threshold payment.

We do not find evidence of differences in treatment effects for the rest of the treatments.

Table B.15: Best linear predictor

	Labor Market		Mental Health	
	Threshold	Piece-rate	Threshold	Piece-rate
	(1)	(2)	(3)	(4)
Mean forest prediction	1.022**	1.162	1.204	0.900***
	(0.451)	(0.963)	(2.220)	(0.238)
Differential forest prediction	1.444*	-5.480	0.727	-4.012
	(0.940)	(2.474)	(1.256)	(1.450)

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. This table reports the best linear predictor for each of the primary treatment cell

Table B.16: ATE on cognitive performance for Mental health- Threshold treatment

Ranking	Estimate	Std- Error
(1)	(2)	(3)
Quartile 1	-0.2996416	0.5305980
Quartile 2	0.2662371	0.5063201
Quartile 3	0.6845878	0.5443749
Quartile 4	1.6134752	0.4741965***

Note: This table reports the average treatment effects of the labor market treatment on cognitive performance under the threshold payment scheme, grouped by the quartiles based on the conditional average treatment estimated using the causal forest algorithm. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.16 displays the different average treatment effects (ATE) for the four CATE rankings. We find that the positive average effect is driven by those in the fourth quartile. Finally, we estimate a differences-in-means model to test whether the prediction for quartiles 2, 3, and 4 is larger than the one in the first quartile. Table B.17 summarizes the results. We find that the only quartiles that significantly differ from each other are quartiles 1 and 4 – while those in the fourth perceived a positive and significant boost from the treatment, those in the first quartile did not. Therefore, we compare the characteristics of these two contrasting groups in Table 1.

Table B.17: Difference-in-means between quartiles

	Estimate	Std. Error	P-value	Adj. P-value
	(1)	(2)	(3)	(4)
Quartile 2 - Quartile 1	0.566	0.735	0.442	0.444
Quartile 3 - Quartile 1	0.984	0.722	0.174	0.295
Quartile 4 - Quartile 1	1.913	0.730	0.009***	0.025***

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. This table reports a difference-in-means estimator, where we test if the predictor for quartiles 2,3 and 4 is statistically different from quartile 1. Column (4) reports Romano & Wolf (2005) adjusted P-values for multiple hypothesis testing with 1,000 replications.