Are Women in Science Less Ambitious than Men? Experimental Evidence on the Role of Gender and STEM in Promotion Applications *

Müge Süer†

December 15, 2023

Abstract

The gender wage gap is to a significant extent driven by gender-based job segregation. One of the potential culprits can be found in supply-side behavioral differences in promotion applications. In this study, using a controlled lab experiment, we disentangle the roles of gender, field of study, and task difficulty in promotion application decisions. Our study provides three crucial findings. First, gender differences in self-limiting promotion application behavior are only present in STEM field students when exposed to a male task. Specifically, when an easier alternative is available, women are less willing to apply for promotions concerning harder tasks than men. Second, there exists no significant difference between men's and women's willingness to apply for promotion concerning female jobs in STEM or non-STEM fields. Third, we find that previously reported gender differences in confidence are present only between STEM field students. The results also suggest that self-sorting into positions does not cause a decrease in overall welfare, however, it causes fewer promotions for women in STEM. We finally propose an easy-to-implement policy intervention to close the gender gap in STEM students when applying for a promotion.

JEL-codes: D91, J16, J62, C91

Keywords: promotion application, self-limiting behavior, hierarchical segregation, STEM, male task, experiment

*We thank Dirk Engelmann, Roel van Veldhuizen, Dorothea Kübler, and the seminar participants at the Berlin School of Economics for helpful comments. Financial support from the German Science Foundation (DFG) through CRC TRR 190 (project number 280092119) is gratefully acknowledged. This study is preregistered as Süer, M. (2022, November 16). Gender, Self-Stereotyping and Job Applications. `osf.io/ra27u`.

†Humboldt-Universität zu Berlin

1 Introduction

Despite progress, the underrepresentation of women in the upper echelons of work remains a prominent issue driving the gender pay gap (Encinas-Martín and Cherian, 2023). This underrepresentation is even more pronounced in science, technology, engineering and mathematics (STEM) fields, which further fuels the gap given the lucrative nature of STEM occupations. The education gap is no longer the primary culprit of the gender pay gap (Blau and Kahn, 2017). Indeed, even with similar levels of numeracy skills, women earn less than their male counterparts (Encinas-Martín and Cherian, 2023). Instead, the gender wage gap is partially explained by occupational segregation (Blau and Kahn, 2017). Occupational segregation manifests itself in two ways, between and within establishments, and it has been shown that especially for college graduates what matters the most is the within-establishment segregation (Barth et al., 2021). We also know that within-establishment segregation is mostly due to promotional differences, which is also called hierarchical segregation in the literature (Bettio et al., 2009). Employers' promotion decisions and employees' application behavior are two sides of the coin leading to this type of segregation. Employer discrimination against women in promotion decisions is a previously documented phenomenon (Blau and DeVaro, 2007; Bohnet et al., 2016; Sarsons, 2017), but it is not the only factor influencing promotion outcomes (Bosquet et al., 2019). Employee behavior, namely the willingness to climb the ladder in work hierarchies, should also be closely examined to develop a better understanding of the female underrepresentation in upper management. In this paper, we scrutinize promotion-related job applications in a controlled environment using two laboratory experiments and focusing on self-limiting employee behavior in this application process that prevents women from reaching advantaged positions in their careers.

Earlier evidence of gender differences in the labor market shows that the career pipeline of women leaks out even before the employer's decision in the hiring process (Clark Blickenstaff, 2005), meaning that women and men themselves may choose different career paths even before an employer makes a decision about their promotion. Promotion decisions are generally considered to be made by the employer, but it has been empirically shown in the French academic setting that application differences account for most of the promotion gap between men and women (Bosquet et al., 2019). Using a series of experiments, Brands and Fernandez-Mateo (2017) highlight the effect of previous negative experiences in the underrepresentation of women in higher management. Additionally, in a stylized online experiment, Exley and Kessler (2022) show that an employee's self-presentation is also a determining factor in how employers make promotion decisions. They demonstrate that a self-presentation gender gap appears in the male-stereotyped domain during the application process. It is important to underline that employees are not automatically promoted just because they deserve it. They are expected to show their deservingness to be promoted or to directly apply for a promotion. Employees must therefore actively participate in the process that will lead them to higher positions in their careers.

In this study, we investigate the self-limiting behavior of individuals in promotion applications by means of two stylized laboratory experiments. We define self-limiting behavior as an individual's personal action that would hinder the individual from reaching their full potential

(Dickerson and Taylor, 2000). Hence, we focus exclusively on employee behavior. To do so, we exploit the potential of lab experiments in three ways. Typical actors of a real-life promotion decision are an employer and the employee. First, to eliminate the previously detected impact of employer expectations on employee behavior (Barbulescu and Bidwell, 2013) and gender differences in self-presentation to the employer (Exley and Kessler, 2022), we have removed the employer component. Second, we have also made job requirements irrelevant by removing them altogether, as Mohr (2014); Coffman et al. (2022); Abraham et al. (2023) demonstrated gender differences in assessing the requirements to apply for a position. Finally, the laboratory setting also helps us separate preferences from self-limiting behavior. An individual's preference not to take on the responsibility of a task is not a self-limiting behavior. Hence, we create a controlled environment in which the self-limiting behavior of men and women in relation to promotion applications is isolated. In our setup, there is no authority or rule to determine deservingness. We start thereby to systematically disentangle reasons behind self-limiting behavior in promotion applications: true performance, past beliefs about absolute and relative performance, future expectations about absolute and relative performance, task difficulty, gender congruency of the task, field congruency of the task and preferences to gain a better understanding of the employee perspective on the promotion gap.

Our study introduces a novel promotion application design, where participants are paired and asked to express their individual willingness to apply for a promotion. The willingness to promote is reported following a real effort task. Only one of the two employees in the group is promoted to the "advanced job" based purely on their relative willingness to apply for promotion. The other employee takes the "novice job". Each of the advanced and the novice jobs provides an individual bonus, the former being greater than the latter. The promoted and unpromoted individuals stay in the same group and work separately in two new real-effort tasks for an additional joint group payoff which is shared equally at the end of the experiment. Both players' efforts contribute to their final earnings. However, the promoted participant's contribution to the joint group account is higher, that is, this promotion involves a shared responsibility, but the promoted employee's output is multiplied by a higher factor. Hence, promoting the better performer is the best strategy for both parties in the group. After the promotion application is completed, both participants perform the tasks.

In our design, there are a total of three promotion application decisions. Across the three decisions, we systematically assign tasks with different difficulty levels to advanced and novice jobs. In the first decision, both jobs have the same low task difficulty level. In the second decision, the task related to the "advanced job" has a higher difficulty level, which makes this particular decision closer to reality. The final decision includes high-difficulty tasks for both jobs. This design allows us to collect three observations for each participant. Thus, we can evaluate self-limiting behavior by exploiting the panel data using a standard difference-in-differences analysis (Card and Krueger, 1994). The second decision, as mentioned above, is the most realistic. But, considering only this decision might raise the question of whether any gender difference in willingness to promote is simply due to a preference for not taking responsibility for the hard task, or whether it is a self-limiting behavior that prevents the employee from applying. To address this, the third option is designed to remove the responsibility effect of the hard task on willingness to promote. If the willingness to promote decreases due to the hard task, then it should also decrease in the third decision. Therefore, we can conduct a difference-in-differences analysis on men and women using the change in task difficulty as the treatment to remove the effect of task difficulty preferences on promotion applications. In addition, the difference between the first and third decisions is analyzed as a difference in willingness to promote when the state of the world consists only of hard tasks.

The study accommodates two separate experiments, Baseline and HighDifficulty. In the first experiment, Baseline, we examine the gender congruency-related differences in behavior and preferences. Gender-incongruency of a task has been previously found to exacerbate gender differences in willingness to take responsibility in group decisions (Coffman, 2014) and self-promotion (Exley and Kessler, 2022). In this experiment, each participant is presented with our experimental design twice in two separate parts, one including a stereotypically masculine and the other a stereotypically feminine task. The order of parts is randomized. Going further, in this experiment we also introduce another congruency level: *field-congruence*. By *field-congruence*, we consider study background and we additionally hypothesize that the gender-incongruency driven differences might shrink when the gender-incongruent task is field-congruent. Therefore, our first experiment distinguishes participants based on their gender as well as their study background, STEM and non-STEM. In the second follow-up experiment, HighDifficulty, we increase the difficulty of the real effort task that they perform before making their promotion choices, therefore implicitly decreasing the ambiguity concerning task difficulty and we concentrate only on the male task and only on STEM students, which was the only combination where we found a gender gap in Baseline. In both experiments, absolute and relative performance-related beliefs are elicited to measure confidence through overestimation and overstatement defined by Moore and Healy (2008). Participants are also asked to state their absolute and relative performance expectations in an easy and a hard task that they would solve in the future. The belief elicitation is not incentivized following Charness et al. (2021) to not cause hedging due to the same experimental design being presented twice with two different tasks in Baseline. The belief elicitation procedure is kept the same in HighDifficulty.

When inspected from Bosquet et al. (2019)'s perspective, each promotion can be regarded as a job application. Focusing on gender differences detected in job application behavior research, we identify three major findings that play a role in the formation of these differences: gender congruency of the domain, competition preferences, and ambiguity. Barbulescu and Bidwell (2013) examine differences in job search and job preferences between men and women using survey and archival data. Their research finds that women are less likely to apply for stereotypically masculine jobs which is partly driven by self-identification and partly by lower expectations regarding a job offer in these domains. Well-crafted laboratory experiments have also shown that gender differences might also depend on the perceived masculinity or the femininity of the task (Coffman, 2014; Dreber et al., 2014; Exley and Kessler, 2022). Exley and Kessler (2022) additionally suggest that there is a self-promotion gap between men and women only in the male domain. Therefore, gender congruency of the domain appears to be an important determining factor behind self selection and we choose to incorporate both male and female domains in our experiment.

Furthermore, experimental evidence from studies conducted by Buser et al. (2014), Flory et al. (2015), and Samek (2019) have identified competitiveness as a crucial factor influencing the career choices of men and women. Women exhibiting a lower preference for competition tend to opt for less prestigious academic tracks, and less competitive work environments, and shy away from competitive compensation schemes. Negative past experiences have also been found to asymmetrically influence women's decisions regarding job applications (Brands and Fernandez-Mateo, 2017), increasing the likelihood that women have less willingness to compete for better positions. Hence, we seek to minimize competition by using piece rates and no limitations for promotion other than self-reported willingness.

On the other hand, ambiguity appears to be another determinant of gender differences in application decisions. Experiments have highlighted the importance of information about the application process. For example, Gee (2019) demonstrated the significant impact of providing information about the number of applicants, which increases the likelihood of women completing their applications. In addition, ambiguity within job requirements is also found to have an effect. A field experiment conducted by Abraham et al. (2023) found that removing redundant job descriptions changed application behavior and that women were particularly sensitive to meeting job requirements, confirming earlier survey evidence (Mohr, 2014). In addition, an online experiment conducted by Coffman et al. (2022) supports the claim that ambiguous job requirements disproportionately affect women, causing them to refrain from applying. Therefore, we have removed any pre-imposed barrier to applying for a promotion in our setup, but the ambiguity related to the difficulty level of the hard task, since our participants are presented only with the easy task before their promotion applications.

The Baseline has three major findings. First, we find that promotion application differences only exist between STEM studying men and women. It is crucial here to underline again that in our study, the promotion application is free of application requirements. Hence, deciding not to apply can only be driven by preferences and beliefs about absolute or relative performance. Our study shows that STEM women have a lower willingness to apply for harder tasks in the presence of an easier alternative. The difference can be explained neither through the belief nor through the preference channel. Second, the promotion application difference becomes insignificant when the female task is used instead. The difference between non-STEM men and women in willingness to apply is on the other hand insignificant for both male and female tasks. Finally, we show that the study background matters and gender differences in absolute and relative confidence in both male and female tasks are mainly driven by STEM-studying males. We also provide evidence for the promotion opportunity loss for women due to self-sorting in jobs.

In our second experiment, HighDifficulty, we use our findings from the Baseline and the research on gender differences in ambiguity aversion regarding gain domain (Eckel and Grossman, 2008; Flory et al., 2015). Moreover, the impact of ambiguity decrease on closing the gender gap has been also demonstrated by Abraham et al. (2023) and Coffman et al. (2022)

in the job descriptions and requirements domain. The BASELINE group demonstrates that statistically significant gender differences in promotion application behavior occur only when a harder task concerns the promoted employee alongside an easier task concerning the non-promoted employee, and only in the male domain. This gender difference is present solely among STEM students. Consequently, HIGHDIFFICULTY targets only STEM students applying for promotions in the male domain. It is important to note that in BASELINE, the first task performed before the promotion decision is the easy task. By contrast, in HIGHDIFFICULTY, participants experience the hard task in the first part of the experiment. This approach reduces ambiguity related to the difficulty level of the hard task, which determines payoffs in relation to promotion decisions. Thus, HIGHDIFFICULTY clarifies for participants the difficulty of the task they must perform when they promote and take responsibility for the group. This intervention successfully eliminates the gender gap in promotion applications, promoting equal labor supply among both men and women.

The rest of this paper is structured as follows: Section 2 explains the detailed designs of our experiment. Section 3 describes our pre-registered hypotheses. Section 4 presents sample characteristics and detailed results. Finally, Section 5 concludes.

2 Experimental Design

The main goal of our study is to disentangle the self-limiting behavior in promotion applications through beliefs and preferences, to check whether this behavior is task or field-of-study dependent, and finally to close this gap by reducing ambiguity related to task content. Our study consists of two experiments, BASELINE and HIGHDIFFICULTY. In BASELINE, we scrutinize gender differences driven by preferences and beliefs in the promotion application behavior with STEM and non-STEM students using both male and female tasks. This experiment is a within-subject design in terms of both tasks being presented to all participants, but inherently a between-subject in terms of study background. After observing the relevance of the field of study on gender differences in BASELINE, we derived the second experiment, HIGHDIFFICULTY, which is a simplified version of our first experiment. In HIGHDIFFICULTY, we propose a policy intervention to close the gap for STEM students in a male task, by reducing the ambiguity concerning the job content, that is the difficulty level. Both of our experiments use the same promotion application design developed for this study.

2.1 Experiment 1: Baseline

Our promotion application design adopted the willingness to contribute concept from Coffman (2014) and was inspired by the design of the laboratory experiment in the online appendix of Coffman et al. (2022). Coffman (2014) uses pairs and asks whether one participant is willing to take responsibility for answering a given question on behalf of the pair. In her experiment, participants decide their willingness to take on this responsibility on a scale of 1 to 4, with 1 being the highest. When the individual responses of the pair are compared, the responsibility for answering the question is given to the person with the highest willingness. Similarly, our design incorporated this mechanism for truthful reporting of willingness to apply for a promotion.

Furthermore, the lab experiment in the Coffman et al. (2022) online appendix attempts to generate a job application setup where participants are asked to choose between a novice and an expert job. The task content of each job was exactly the same, but the jobs differed in terms of pay. The reason for this design choice of having the same task for both jobs has been reported as eliminating the effect of preferences. Therefore, the authors have argued that they were able to focus on beliefs.

Our first experiment, Baseline, consisted of two main parts that were identical except that they used different real effort tasks based on two comparable modules of the Paper-based TestAS[1], a test used for the university admissions of international students to German universities. The first part comprised questions from the Solving Quantitative Problems module of TestAS as the main effort task. The other part of the experiment used the Recognizing Linguistic Structures module of the same test. These two modules were chosen to address the gender congruency and their effects on behavior. The modules were also comparable regarding the necessary time per question. Using Paper-based TestAS also allowed us to control for task difficulty. The questions of TestAS were categorized into three pre-determined difficulty levels easy, medium, and hard, which gave the flexibility to credibly alter the degree of difficulty. The order of these two parts was randomized.

In each part of Baseline, there were two sections. Following an introduction, subjects proceeded to the first section of that part. The first section consisted of six questions from the relevant module with two difficulty levels, i.e. three low-difficulty and three medium-difficulty questions. Participants were given ten minutes to answer these questions. For each correctly solved question in this section, participants earned 1 Euro as a reward. Once participants had completed the questions, they were then asked to report their absolute and relative performance expectations regarding their performance on these questions. In addition to their expectations about the past, they were also asked to state their future expectations. These future expectations were elicited for one similar and one more difficult task. None of the belief elicitations was incentivized following the findings of Charness et al. (2021). This approach was taken to prevent any potential performance manipulation through hedging in the second block, as we adopted a within-subject design.

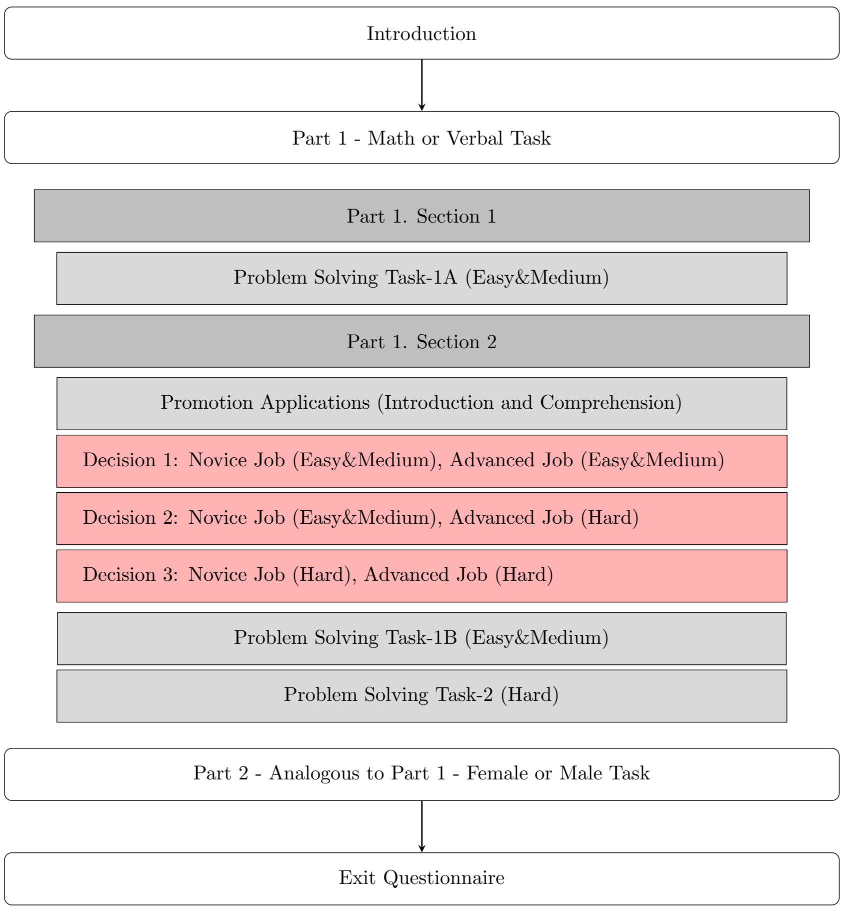
In the second section of each part, participants were presented with the task of making three independent stylized promotion application decisions. The section began with an explanation of the application procedure, followed by a comprehension question to ensure understanding. A promotion application decision entailed expressing the willingness to take responsibility for an advanced job within a two-person group. Prior to making the decisions, participants were informed that they would be matched with a random person from the lab and no personal information about this individual would be disclosed. In this section of the experiment, the pair had a chance to generate additional bonuses together by correctly answering new questions in the

[1]Detailed information related to Paper-based TestAS can be found in the following file: https://www.testas.de/en/modellaufgabenpapier_en.htm

same category they were previously tested on, i.e. Solving Quantitative Problems or Recognizing Linguistic Structures (Please refer to Figure 1 for the structure of each experimental session). After the promotion decisions were made, participants completed the new questions that were the subject of their promotion. No performance feedback was given in any of the sections.

The experiment then continued with the second part. This part was analogous to the first, the only difference being the task type. The task type to start with was randomized due to the within-subject design feature of the experiment. The experiment ended with an exit questionnaire where participants were asked to answer demographic questions along with a single-item risk preference question used in the German Socio-Economic Panel (SOEP) survey (Goebel et al., 2019), single-item competition preferences question by Fallucchi et al. (2020), Big 5 personality traits, continuous gender identity measure by Brenøe et al. (2022), family education background, migration background, past exposure to TestAS, views on performance tasks, and reliability. Participants were also asked to report the stereotype associated with the task type.

Figure 1: Experimental Flow of Baseline



Decisions on promotion applications were designed so that there were two different roles in the group. We called these the advanced job and the novice job. Participants were asked to indicate their willingness to apply for the advanced job for their group. In each of the three decisions, only one member of the group was hired for the advanced job, based on their reported willingness to apply. The other member was then assigned to the novice job. The advanced job

Table 1: Expected Earnings Based on Expected Rank and Expected Number of Correct Answers

Expected Rank →	Below Median		Above Median	
Expected Correct Answers ↓	Novice Job	Advanced Job	Novice Job	Advanced Job
0	7.00	5.25	2.00	4.00
1	8.50	7.50	3.50	6.25
2	10.00	9.75	5.00	8.50
3	11.50	12.00	6.50	10.75
4	12.00	14.00	8.00	13.00

Notes: The figures exclude show-up fees and are in EUR.

offered a higher earning potential with a 4 Euros individual promotion bonus. It contributed 4 Euros per correct answer to the group earnings which was shared equally between the promoted and the nonpromoted participants of the group. On the other hand, the novice job contributed 1 Euro per correct answer to the group earnings and it offered an individual novice bonus of 2 Euros. Table 1 reports the expected earnings based on individuals' absolute and relative performance expectations in this mechanism. It reveals that for a risk-neutral agent stating the highest willingness to apply for the advanced job is the dominant strategy if the person thinks that their performance is above the median. If the performance expectations are below the median but the median is three correct answers or above, stating the highest willingness to apply should also be preferred. Thereby, the incentive scheme of the stylized promotion applications motivated participants towards selecting the advanced job as long as they thought they were good enough.

Participants viewed one decision at a time on the screen and chose a number between 1 and 4 to represent their willingness to take the advanced job for their group. A rating of 1 indicated the highest willingness, while a rating of 4 indicated the lowest willingness. If both chose the same number, the computer flipped a coin to determine who got the advanced job within the group. Framing the reporting as a 1 to 4 willingness representation gave us the chance to collect a richer measure with higher variance. This framing was also important to reflect real-life promotion applications since willingness to be promoted is not always a binary decision but is actually about showing how willing one is to get promoted.

The three application decisions were designed to separate the effect of their future performance belief in similar tasks, future performance belief in harder tasks, and their preference for signing up for a harder task in the presence of an easy alternative. Therefore, the difficulty level of the jobs that were subject to the advanced job and the novice job was systematically altered. In the first decision (hereafter *Easy-Easy*), both the novice and advanced jobs concerned two low, and two medium-difficulty (four in total) questions, meaning the difficulty level was the same as their first section but it had new questions. The third decision (hereafter *Hard-Hard*) had also the same task for both jobs, but it involved solving four high-difficulty questions. Participants were informed that the difficulty level of the first section was easy and medium. So they had a clear idea of what to expect in terms of difficulty when presented with a low and medium-difficulty task. However, for the hard questions, they were simply told that the questions were from the same domain but were more difficult. This made the hard-difficulty task relatively unfamiliar and ambiguous and required participants to make decisions based on the expectations of their performance in a task with higher difficulty. The second decision (hereafter *Easy-Hard*) presented a choice between two low and two medium-difficulty (four in total) questions for the novice job and four high-difficulty questions for the advanced job. Therefore, the second decision required participants to weigh their preference for taking on a more challenging task against the option of an easy but less important position in the group. In the end, one of the three decisions was randomly picked and the participant's earnings in the second section were calculated based on the job allocation and group performance subject to this decision.

Following application decisions, subjects performed both easy-medium and hard tasks without learning the results of their applications. They had six minutes to solve four easy-medium questions and eight to solve four hard questions. The outcomes of their performance on these tasks, combined with the randomly chosen application decision, determined their payoffs. This design feature allowed us to generate true performance for both the easy-medium task and the hard task since participants did not know which one of the decisions was picked or whether they were assigned the advanced job responsibility. After each task, we also elicited their absolute and relative performance beliefs.

It is important to remember that the within-subject design of our experiment allowed us to create panel data for each participant including all of their decisions and performances concerning both the male and the female tasks. Therefore, thanks to this design, we had the chance to truly follow each person and scrutinize the changes in their decisions and beliefs when the task type and difficulty changed.

In Baseline, we recruited 303 participants from two active social science labs in Berlin. The experiment was run at the WZB-TU lab of the Technical University of Berlin and the PLEX lab of the University of Potsdam.[2] The Baseline experiment took an average of 75 minutes. It was run between November 16, 2022, and March 15, 2023, in both labs. The average earnings before the show-up fee are 10.08 Euros for the TU lab and 9.74 Euros for the Potsdam lab. The earnings difference between labs is not significant (Wilcoxon rank-sum test, $z = -1.048$, $p = 0.2948$). The show-up fees were calculated based on lab-specific regulations, the TU lab show-up fee being 6 Euros and the Potsdam lab 9 Euros.

2.2 Experiment 2: HighDifficulty

In our second experiment, HighDifficulty, we focused exclusively on STEM students and used only the math task to examine the policy intervention. The reason we focus on this group is due to the gender gap found in the first experiment, which will be discussed in more detail later in Section 4. Briefly, in our first experiment, we find that STEM women tend to have a lower willingness to apply for a promotion when the novice-level job is an easy task and the advanced job is a hard task, compared to STEM men. This finding tells us that STEM women prefer easier jobs when they are available.

[2] The PLEX lab of the University of Potsdam is later added to the experiment due to the lack of non-STEM students in the WZB-TU lab.

We recruited our sample ($N = 158$) only from the WZB-TU Laboratory of the Technical University of Berlin since the study only included a comparison of STEM students. HighDifficulty consisted of only one part and two sections since we wanted to examine gender differences in the math task following our Baseline results. In this experiment, we presented the new participants with hard-difficulty questions instead of the low and medium-difficulty questions used in the first section of Baseline. The purpose of this change was to minimize ambiguity about the level of difficulty that would be later influential in the promotion decisions that involved hard difficulty tasks. Due to the increased difficulty, the time allotted for the first section of HighDifficulty was extended to twelve minutes instead of ten. The rest of the experiment remained exactly the same.

3 Hypotheses

The study aims to test five sets of hypotheses through two separate lab experiments. The first four of these are addressed in our first experiment, Baseline. The fifth is addressed in HighDifficulty.[3]

3.1 Experiment 1: Baseline

Our first set of hypotheses focuses on beliefs. We begin by focusing on the influence of educational background on beliefs about past performance. We propose that the field of study plays a crucial role in shaping both absolute and relative performance expectations. In our first experiment, we employed two different tasks, math and verbal, based on their gender congruency, and we recruited both STEM and non-STEM subjects. Thus, this design allows us to examine the effect of field congruence on performance expectations. By systematically manipulating these factors, we aim to examine how beliefs are influenced and potentially altered by the congruence between the task domain and the participant's field of study. We hypothesize that the field of study serves as a relevant factor in the formation of performance-related beliefs.

Hypothesis 1.A (Beliefs)**.** *The beliefs about past performance in different tasks are driven by congruency with the field of study.*

In Subhypothesis 1.A, we propose that task congruence with the field of study might predict performance-related beliefs. Early studies focusing on gender differences in confidence find men to be more confident Croson and Gneezy (2009). Elaborating further on this finding, later studies have found that men are generally more confident specifically in male tasks, meaning tasks in which men are expected to perform better than women (Coffman, 2014; Dreber et al., 2014; Coffman et al., 2022; Exley and Kessler, 2022). A recent finding scrutinizes gender differences by taking the study track into account and suggests that the gender gap in reported self-assessment exists only for students in the science track in math, science, and social science (Saygin and

[3]Our hypotheses are all preregistered. They are regrouped for better readability without changing the content. The hypothesis for the second experiment was preregistered after the first experiment was run, in an additional document under the same repository.

Atwater, 2021). They found no gender differences in self-assessment in non-science fields. Saygin and Atwater (2021) also show that women report better self-assessment in literature, a very female domain, regardless of their field of study. It is important to emphasize that reported self-assessment, which is a rather subjective measure, is not necessarily the same as performance expectations. Therefore, in Subhypothesis 1.B, we start from the more general finding that males are more confident than females (Croson and Gneezy, 2009) and test whether males have higher average absolute and relative levels of confidence compared to females in the same field of study.

Hypothesis 1.B (Beliefs)**.** *Given the field of study, men, on average, are expected to be more confident than women in all past and future performance expectations.*

In our study, we additionally speak to the concept of imposter syndrome, initially defined as "*an internal experience of intellectual phoniness*" by Clance and Imes (1978). It is important to note that both the methodology for detecting imposter syndrome and the phenomenon itself have not been extensively studied empirically yet (Whitman and Shanine, 2012). Therefore, in this study, we concentrate on the differences between past and future performance expectations. We examine the phenomenon by analyzing the disparity between future and past performance expectations in the initially performed real effort task in the first session of each part. Thereby, we define imposter syndrome as beliefs about future performance in a similar task being worse than beliefs about past performance. Building on previous evidence suggesting imposter syndrome being more prevalent in women (Clance and Imes, 1978; Chrousos and Mentis, 2020), we propose in Subhypothesis 1.C that women would be suffering from imposter syndrome more than men. Moreover, in the successful identification of the phenomenon, we expect the level of the imposter syndrome to be lower when the performed task is congruent with the field of study.

Hypothesis 1.C (Beliefs)**.** *Imposter syndrome is, on average, more prominent in women. It can be partially explained by task congruency with the field of study.*

In the second hypothesis set, we focus on understanding the differences between men and women in terms of their willingness to apply for promotions in gender or field-congruent domains. We build this set of hypotheses by combining the literature about job applications in general with the Bosquet et al. (2019) perspective regarding promotion applications as job applications. We propose three hypotheses related to the differences in willingness to apply for promotions. First, drawing from the findings of studies conducted by Coffman et al. (2022) and Exley and Kessler (2022), we hypothesize that women, overall, exhibit a lower willingness to apply compared to men in the male domain. Second, we do not expect to find gender differences in the female domain.

Hypothesis 2.A (Willingness to Apply for Promotion)**.** *Women, compared to men, have a lower willingness to apply for promotion for the male task.*

Hypothesis 2.B (Willingness to Apply for Promotion)**.** *There is no gender difference in willingness to apply for promotion for the female task.*

Next, we examine the role of field congruency in driving these differences. We expect that the disparity in willingness to apply is particularly pronounced among women in non-field congruent domains, specifically non-STEM fields. We assume that self-reliance and individual preferences may play a role in shaping these differences, as women in non-congruent fields might feel even less comfortable than women in STEM fields when it comes to male-dominated domains.

Hypothesis 2.C (Willingness to Apply for Promotion)**.** *The difference in willingness to apply for promotion is driven by women in non-STEM fields.*

We propose two hypotheses related to task difficulty preferences in the context of gender and field congruency. Based on the previous literature on difficulty preferences (Gneezy et al., 2003), Hypothesis 3.A suggests that women, compared to men, have a preference for easier tasks in the male task. This hypothesis posits that women tend to lean towards tasks that are perceived as less challenging or demanding.

Hypothesis 3.A (Task Difficulty)**.** *Women, compared to men, prefer easier tasks in the male-type domain.*

Hypothesis 3.B states that the expected difference in task difficulty preferences between men and women, as outlined in Hypothesis 3.A, decreases when the task is congruent with the field of study. In other words, when the task aligns with the individual's field of study, the gender-based preference for easier tasks becomes less pronounced. This hypothesis implies that field congruency has a moderating effect on the task difficulty preferences of men and women. It is motivated by the general positive selection expectations about women who are already in the STEM field (Saygin and Atwater, 2021).

Hypothesis 3.B (Task Difficulty)**.** *The gender difference in preferences for an easier task in the male domain is weaker when the task is congruent with the field, i.e. among STEM students.*

Coffman (2014) calculates financial losses due to women who are not willing to contribute their ideas. In this study, despite their correct answers, women report a lower willingness to contribute their ideas to their group, resulting in losses to the group account. Therefore, we hypothesize that the inefficient sorting of individuals into jobs due to the gender difference in willingness to apply for a promotion results in a significant loss in total earnings. This hypothesis suggests that when individuals are not appropriately matched to promotions based on their deservingness, it leads to suboptimal outcomes in terms of total earnings.

Hypothesis 4. *The inefficient sorting to jobs causes a significant loss in total earnings.*

3.2 Experiment 2: HighDifficulty

In line with the literature documenting gender differences in ambiguity aversion (Eckel and Grossman, 2008; Flory et al., 2015), we propose that women's higher preference for easier tasks might be due to ambiguity concerning task difficulty. Therefore, we hypothesize that presenting participants in the STEM field with a hard male task diminishes the gender difference detected in Baseline. This hypothesis suggests that reducing ambiguity can play a crucial role in mitigating gender differences in willingness to apply for promotion within the STEM domain.

Hypothesis 5. *In the STEM field, presenting participants with a hard task first reduces the gender difference in the preference for easier tasks when they are available.*

4 Results

4.1 Data and Summary Statistics

Our first experiment, Baseline, consists of 303 and our second experiment, HighDifficulty, of 158 participants. The mean age of the pooled data is 25 with a median of 24. The entire sample is 62.5% German. 10.2% of the total sample has been exposed to TestAS before, and finally, the median of our sample is currently enrolled in a bachelor's program. The subgroup sample sizes and descriptives can be found in table 2.

The performance difference in math questions between STEM men and STEM women in Baseline is not statistically significant (Wilcoxon rank-sum test, $z = 1.388$, $p = 0.1651$). However, it is significant amongst non-STEM students at a 5% level (Wilcoxon rank-sum test, $z = 2.103$, $p = 0.0355$). Additionally, the math performance difference amongst STEM students in HighDifficulty is also not significant (Wilcoxon rank-sum test, $z = 0.231$, $p = 0.8170$). The decrease in overall math performance for STEM students from Baseline to HighDifficulty can be attributed to the increased difficulty of the task in the first section (Wilcoxon rank-sum test, $z = 12.568$, $p < 0.0000$).

There are no significant performance differences in the verbal task of the first section between STEM men and women, and non-STEM men and women in Baseline (Wilcoxon rank-sum test, $z = -0.694$, $p = 0.4879$ and $z = -0.154$, $p = 0.8777$, respectively).

When it comes to willingness to apply for a promotion, in the math task of Baseline, we see that the only statistically significant difference appears in Decision 2 between STEM men and STEM women (Wilcoxon rank-sum test, $z = -2.801$, $p = 0.0051$). The difference between non-STEM men and non-STEM women is borderline significant at 10% level (Wilcoxon rank-sum test, $z = -1.659$, $p = 0.0970$). The difference between STEM men and STEM women becomes insignificant in the math task in HighDifficulty (Wilcoxon rank-sum test, $z = -0.574$, $p = 0.5657$). We find no statistically significant difference when we compare men and women in any of the decisions in the verbal task.

4.2 Beliefs

The Field of Study is a Significant Predictor of Past Performance Beliefs

As a first result, we delve into field congruency and its relationship to past performance beliefs. We run a regression pooling all STEM and non-STEM students together, controlling for field congruency, task type and their interaction, and actual performance. As seen in the first two columns of table 3, field congruency does seem to matter only for the math task when we pool

Table 2: Summary Statistics

	Baseline				HighDifficulty	
	STEM		Non-STEM		STEM	
	Male	Female	Male	Female	Male	Female
Sample Size	80	73	70	80	85	73
Age	25.2	24.5	25.1	25.2	25.3	25.0
	(5.1)	(4.4)	(7.6)	(7.2)	(4.8)	(4.4)
Median Age	24	24	23	23	24	24
German %	62.5%	49.3%	82.9%	72.5%	61.2%	46.6%
Previous TestAS Exposure %	6.2%	5.5%	11.4%	8.8%	15.3%	13.7%
Median Current Education	Bachelor	Bachelor	Bachelor	Bachelor	Bachelor	Bachelor
Performance (Math) - Sec 1	4.26	4.01	3.97	3.53	1.95	1.96
Out of 6 questions	(0.95)	(1.07)	(1.18)	(1.27)	(1.09)	(1.30)
Willingness to Apply (Math)						
Decision 1	1.33	1.51	1.86	1.91	1.79	1.88
Easy-Easy	(0.63)	(0.80)	(1.06)	(1.10)	(0.94)	(0.96)
Decision 2	2.00	2.51	2.60	2.85	2.58	2.66
Easy-Hard	(0.95)	(1.13)	(1.00)	(1.15)	(1.03)	(1.15)
Decision 3	1.65	1.79	2.16	2.25	2.01	2.02
Hard-Hard	(0.97)	(1.04)	(1.21)	(1.19)	(1.07)	(1.14)
Performance (Verbal) - Sec 1	3.43	3.55	3.10	3.19		
Out of 6 questions	(1.27)	(1.38)	(1.51)	(1.38)		
Willingness to Apply (Verbal)						
Decision 1	1.63	1.53	1.57	1.68		
Easy-Easy	(0.96)	(0.85)	(0.89)	(0.87)		
Decision 2	2.10	2.38	2.54	2.63		
Easy-Hard	(1.01)	(1.11)	(1.05)	(1.02)		
Decision 3	1.81	1.78	1.80	1.75		
Hard-Hard	(1.07)	(1.06)	(1.04)	(0.97)		

Notes: Unless stated otherwise, figures without parentheses represent means and those within parentheses are standard deviations.

the data. We see that confidence in the math task is lower overall, but higher when the students are in the STEM field. However, when we run separate regressions we see in columns three and four that the non-STEM students have significantly higher absolute and relative performance expectations in the verbal task than in the math task. Therefore, we conclude that the field of study is an important element in beliefs about past performance.

Table 3: Pooled Performance Expectations in Baseline

	All		Non-STEM		STEM	
	(Absolute)	(Relative)	(Absolute)	(Relative)	(Absolute)	(Relative)
Field Congruent (FC)	-0.133	-0.280	0.341*	0.716***	0.0218	-0.184
	(0.143)	(0.161)	(0.138)	(0.148)	(0.133)	(0.131)
Math Task	-0.471**	-1.003***				
	(0.145)	(0.166)				
FC × Math Task	0.637**	1.123***				
	(0.226)	(0.271)				
Correct Answers	0.258***	0.233***	0.263***	0.221***	0.274***	0.269***
	(0.0374)	(0.0418)	(0.0555)	(0.0621)	(0.0540)	(0.0595)
Controls	YES	YES	YES	YES	YES	YES
Obs	606	606	300	300	306	306

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Random effects model with robust standard errors. The dependent variables are pooled past performance expectations about absolute and relative performances. Controls include age, education, employment status, German origins, session fixed effects and task order variables.

STEM Men are More Confident

In our Baseline, we find that STEM males are significantly more confident than STEM females in the math task, the male domain. On the other hand, as shown in the third column of Table 4, non-STEM males expect significantly higher absolute performance than non-STEM females on the math task, which is consistent with previous findings. Again, replicating previous findings in the literature, we find no gender differences in the verbal task between STEM males and females or between non-STEM males and females. When we pool the sample over males and females without splitting it based on the field of study, we replicate the previous findings suggesting men are more confident in the male domain in terms of both absolute and relative performance and female tasks do not reveal any gender differences (See Appendix A.1.1 Table 13).

The interesting part of this exercise begins when we compare different groups. First, we compare STEM to non-STEM males and STEM to non-STEM females. In the comparison of STEM to non-STEM males, we see that STEM males are significantly more confident than their non-STEM counterparts in the female domain, the verbal task (see Appendix A.1.1 Table 14). They also report higher relative past performance expectations in the male domain. However, we do not find a difference between STEM and non-STEM females in any of the beliefs. Second, when we compare non-STEM males to STEM females, we also find no significant difference in either task (see Table 5). These results provide the first evidence of overconfidence among

Table 4: Past Performance Expectations in BASELINE by Field of Study

	Math				Verbal			
	STEM		Non-STEM		STEM		Non-STEM	
	(Abs)	(Rel)	(Abs)	(Rel)	(Abs)	(Rel)	(Abs)	(Rel)
Male	0.481**	0.497*	0.499*	0.191	0.274	0.103	0.0976	0.230
	(0.175)	(0.211)	(0.220)	(0.235)	(0.187)	(0.209)	(0.219)	(0.258)
Performance	0.201*	0.312**	0.238*	0.227*	0.166*	0.0912	0.255**	0.199*
	(0.0821)	(0.0980)	(0.0925)	(0.105)	(0.0754)	(0.0745)	(0.0773)	(0.0878)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Obs	153	153	150	150	153	153	150	150
R^2	0.430	0.389	0.353	0.330	0.299	0.254	0.304	0.257

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS model with robust standard errors. The dependent variables are elicited past performance expectations about absolute and relative performances. Controls include age, education, employment status, German origins, session fixed effects and task order variables.

STEM men relative to women and their non-STEM male counterparts. Finally, if we compare STEM men to non-STEM women, we see how gender differences are inflated, how differences in math become highly significant at the 0.1% level and how relative performance expectations in verbal also become significant (see Appendix A.1.1 Table 15). We, therefore, conclude that the field of study is an essential piece of information to provide more conclusive analyses of gender differences in confidence. If we do not control for the field of study composition of our sample, we tend to overestimate or underestimate differences.

Table 5: Past Performance Expectations in BASELINE Comparing Non-STEM Men to STEM Women

	Math		Verbal	
	(Abs)	(Rel)	(Abs)	(Rel)
Male	0.127	-0.110	0.00197	-0.0469
	(0.276)	(0.281)	(0.263)	(0.272)
Performance	0.226**	0.455***	0.123	0.0457
	(0.0843)	(0.0945)	(0.0801)	(0.0765)
Controls	YES	YES	YES	YES
Obs	143	143	143	143
R^2	0.362	0.433	0.319	0.319

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS model with robust standard errors. The dependent variables are elicited past performance expectations about absolute and relative performances. Controls include age, education, employment status, German origins, session fixed effects and task order variables.

As mentioned above, the performance-related beliefs were elicited for both past and future expectations. In addition, future expectations were also categorized as future expectations in a familiar (easy and medium difficulty in BASELINE) and in a harder task. Thus, we run the same regressions for the past performance expectations for both types of future expectations (See Appendix A.1.2 and A.1.3 for familiar and harder tasks respectively). The results show that STEM men are also more confident about their future performance in a harder task. The slight overconfidence of non-STEM men compared to non-STEM women in absolute performance expectations in the math task disappears when we ask about their future expectations. The results for STEM men compared to non-STEM men remain the same for the relative performance expectations in both the math and verbal tasks. There is no difference between non-STEM men and STEM women in either task. Finally, the gender differences are again inflated when we compare STEM men to non-STEM women. We find the same results in the future expectations in a similar task (See Appendix A.1.2). Therefore, our study provides clear evidence of STEM men's overconfidence in all types of elicited beliefs, except in the female domain when compared to STEM women. Moreover, we find no evidence of imposter syndrome as we define it.

4.3 Willingness to Apply for Promotion

STEM Men Have the Highest Willingness to Apply for Promotion in the Male Task

In our preregistered hypotheses, we have stated that women, compared to men, would have a lower willingness to apply in the male-type task and that the difference would be larger for the non-STEM women. On the other hand, we have not expected gender differences in willingness to apply in the female task following the findings of Exley and Kessler (2022).

Table 6 contains two specifications of the pooled regressions of willingness to apply. This analysis aims to understand the general tendency to apply for the advanced job, independent of task difficulty. Therefore, we pool all three observations per participant in each task type and run a random effects regression on the *willingness to apply* for a promotion variable.[4] Specification (1) of the table runs the regression for the math task and (2) for the verbal task. If we structure the regression using STEM men as a reference point, we see that all other groups have a significantly lower willingness to apply than STEM men when the task is male. The largest difference is again between STEM men and non-STEM women, as in the analysis of beliefs. If we compare non-STEM men with non-STEM women or non-STEM men with STEM women, we find no gender differences in the male task (see Appendix A.2.1 Table 22 Columns 1 and 3). In the female domain, no difference is found as expected, however, non-STEM males are found to be borderline significantly more willing to apply for promotion in the female domain at 10% level, which is a new finding that has not been extensively explored in the literature (see Appendix A.2.1 Table 22 Columns 2 and 4).

4.4 Task Difficulty

Our design's feature of generating panel data, specifically six data points per participant, makes it possible to compare different promotion decisions made by the same participant. This allows us to use the standard difference-in-differences (DiD) approach of Card and Krueger (1994) for inferring causality in experimental gender studies.

[4]It is important to recall at this point that the *willingness to apply* variable is not binary. It is reported on a scale of 1 to 4, with 1 being the highest. Therefore, in this table, a lower number means a higher willingness.

Table 6: Pooled Willingness to Apply

	Math (1)	Verbal (2)
Non-STEM Males	0.467***	0.00383
	(0.141)	(0.152)
STEM Females	0.340**	0.126
	(0.131)	(0.141)
Non-STEM Females	0.608***	0.131
	(0.137)	(0.135)
Correct Answers	-0.0800	-0.109***
	(0.0419)	(0.0316)
Controls	YES	YES
Observations	909	909

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: GLS random effects model with robust standard errors. The dependent variable is the willingness to apply pooling all three decisions per task type. Controls include age, education, employment status, German origins, session fixed effects and task order variables.

Only STEM Women Prefer Easier Math Tasks When Available

We run our first DiD analysis between the second, *Easy-Hard*, and third, *Hard-Hard*, decisions of the application section. The advanced job in *Easy-Hard* has a hard task, while the novice job has a familiar easy-medium task. On the other hand, *Hard-Hard* has a hard task in both jobs. The interaction term between gender and the second decision (Female x *Easy-Hard*) in our random effects model shows that compared to men, women prefer easier tasks only in the male-type task and only among STEM students (see table 7). Surprisingly, the effect is smaller but still borderline significant for the female task when we compare STEM men and women. There is no gender difference between non-STEM students in any of the tasks.

Table 7: Difference in Differences between *Easy-Hard* and *Hard-Hard* Decisions

	STEM		Non-STEM	
	Math	Verbal	Math	Verbal
Female	0.185	0.0653	0.152	0.141
	(0.178)	(0.185)	(0.196)	(0.177)
Easy-Hard	0.350**	0.288*	0.443**	0.743***
	(0.122)	(0.116)	(0.152)	(0.146)
Female x *Easy-Hard*	0.362*	0.315	0.157	0.132
	(0.180)	(0.179)	(0.211)	(0.197)
Performance	-0.0890	-0.133*	-0.0920	-0.0821
	(0.0768)	(0.0616)	(0.0599)	(0.0468)
Controls	YES	YES	YES	YES
Observations	306	306	300	300

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: GLS random-effect model with robust standard errors. The dependent variable is the willingness to apply, pooling the second and third decisions. Controls include age, education, employment status, German origins, session fixed effects and task order variables.

Now we briefly look at the hypothetical case where the DiD approach was not implemented. Table 8 shows the standard OLS regression on the *Easy-Hard* decision. It is noteworthy that the female coefficient is relatively inflated in this regression. Furthermore, solely examining these decisions would not address whether this gender difference is due to a general dislike of the hard task that ultimately affects the promotion application, or whether it is due to a preference for not taking responsibility for the hard task when there is an easier alternative to choose. The coefficient of *Female* in Table 7 shows that women do not in fact dislike a hard task, they take the responsibility of the group not significantly different from men when both jobs have the hard task in the *Hard-Hard* decision. However, in this case, women have a stronger preference than men for the easy and more familiar task when it is available.

Table 8: Simple OLS on *Easy-Hard* Decision

	STEM		Non-STEM	
	Math	Verbal	Math	Verbal
Female	0.631**	0.467*	0.195	0.198
	(0.189)	(0.198)	(0.208)	(0.197)
Performance	-0.0729	-0.0612	-0.168*	-0.0154
	(0.0906)	(0.0752)	(0.0776)	(0.0578)
Controls	YES	YES	YES	YES
Observations	153	153	150	150

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS model with robust standard errors. The dependent variable is the willingness to apply in the second decision, *Easy-Hard*. Controls include age, education, employment status, German origins, session fixed effects and task order variables.

As a next DiD exercise, we compare the first, *Easy-Easy*, to the third, *Hard-Hard*, promotion application decision. In the first decision, both advanced and novice jobs have an easy-medium task that is similar to the first section of the relevant part. Table 9 shows that there is no gender difference in willingness to apply when the promotion application decision is between the same task options. Independent of their difficulty, we do not find gender differences when both the promotion and the novice tasks are the same, showing us that STEM women are less willing to apply for a promotion when the promotion task is hard, only if the easier and more familiar alternative is available.

Sample Composition Matters

These two DiD analyses tell us two key things. First, women are not necessarily different from men when it comes to taking responsibility for a difficult task. However, STEM women do when an easier alternative is available. Second, we see that non-STEM women, compared to their non-STEM male counterparts, are overall less willing to be promoted in their verbal female domain, which contradicts the self-stereotyping behavior revealed in the previous literature. Taken together, these findings lead us to reject our task difficulty-related hypotheses where the literature had motivated us to expect that women would prefer easier tasks and where we had expected that field congruency might have compensated for gender differences.

Table 9: Difference in Differences between *Easy-Easy* and *Hard-Hard* Decisions

	STEM		Non-STEM	
	Math	Verbal	Math	Verbal
Female	0.121	0.00141	0.212	0.176
	(0.179)	(0.185)	(0.195)	(0.176)
Easy-Easy	-0.325**	-0.188*	-0.271*	-0.229*
	(0.101)	(0.0922)	(0.122)	(0.0896)
Female x *Easy-Easy*	0.0373	-0.0591	-0.0661	0.154
	(0.137)	(0.158)	(0.173)	(0.129)
Performance	-0.0953	-0.133*	-0.0182	-0.150**
	(0.0696)	(0.0559)	(0.0675)	(0.0533)
Controls	YES	YES	YES	YES
Observations	306	306	300	300

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: GLS random-effect model with robust standard errors. The dependent variable is the willingness to apply, pooling the first and third decisions. Controls include age, education, employment status, German origins, session fixed effects and task order variables.

Notably, our results show that the gender differences are mainly in the areas where the participants are more proficient. So we do another analysis with STEM men and non-STEM men, repeating the same exercises from Table 7 and 9. We can see from this table in our Appendix A.2.2 that if we had not balanced our recruitment in terms of field of study and had inadvertently recruited an unbalanced sample of STEM men and non-STEM women, our results would have been more in line with the previous literature, with women overall being less willing to apply on the male task and with gender differences being found in the female domain, where women would not prefer the hard task.

The Gender Gap between STEM Students in Preferring Easier Math Task is Partially Belief Driven

Based on our belief elicitation questions, we can explore both past performance-related beliefs and future performance-related beliefs. We now try to understand the reasons behind the differences in willingness to be promoted between STEM men and women in the math task. Table 10 shows that beliefs do matter for reported willingness to apply for promotion. Yet, beliefs about past performance explain less of the difference than beliefs about future performance. As mentioned above, our participants in Baseline were asked to report their absolute and relative performance expectations on a future task that would be harder than the one they had originally performed. Thus, we see that future beliefs (both about absolute and relative performance) can explain women's preferences for easy task when it is available slightly better than beliefs about past performance. However, although the gap is no longer significant, it does not close. Thus, we now know that future performance expectations matter, but are not enough to close the gap. We have therefore hypothesized that this difference might be related to ambiguity, that is a lack of knowledge about the difficulty level of a future hard task in our study, and preregistered our second experiment, HighDifficulty.

Table 10: Difference in Differences between *Easy-Hard* and *Hard-Hard* Decisions in Baseline

	STEM in Math				
Female	0.185	0.135	0.0912	-0.108	-0.133
	(0.178)	(0.186)	(0.185)	(0.186)	(0.187)
Easy-Hard	0.350**	0.773*	0.936**	0.284	0.291
	(0.122)	(0.389)	(0.363)	(0.313)	(0.352)
Female x *Easy-Hard*	0.362*	0.317	0.303	0.247	0.250
	(0.180)	(0.185)	(0.183)	(0.210)	(0.206)
Performance	-0.0890	-0.0585	-0.0115	-0.0332	-0.0746
	(0.0768)	(0.0773)	(0.0763)	(0.0811)	(0.0696)
Abs Past Performance		-0.104			
		(0.0894)			
Abs Past x *Easy-Hard*		-0.0947			
		(0.0809)			
Rel Past Performance			-0.187*		
			(0.0828)		
Rel Past x *Easy-Hard*			-0.122		
			(0.0668)		
Abs Future Performance				-0.303**	
				(0.0940)	
Abs Future x *Easy-Hard*				0.0591	
				(0.0781)	
Rel Future Performance					-0.252***
					(0.0687)
Rel Future x *Easy-Hard*					0.0475
					(0.0776)
Controls	YES	YES	YES	YES	YES
Observations	306	306	306	282	282

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: GLS random-effect model with robust standard errors. The dependent variable is the willingness to apply, pooling the second and third decisions. Controls include age, education, employment status, German origins, session fixed effects and task order variables.

4.5 Inefficiency

Self-selection into Promotions Does Not Reduce Average Earnings, but It Does Reduce the Number of Women Promoted

Moving forward, we also examine whether there is a loss of earnings due to inefficient sorting into promotion positions. Assuming that participants base their decisions solely on their true relative past performance within their pairings, we expect that their total earnings should not differ from earnings based on their willingness to apply. Our analysis focuses on STEM students for the math task and non-STEM students for the verbal task. Table 11 indicates that there is no significant decrease in the average group bonus between willingness-based sorting and past performance-based sorting, except for the first decision in the verbal domain among non-STEM students. This analysis indicates that participants have a strong understanding of their previous performance overall. Thus, no loss is found when comparing earnings generated by willingness-based sorting with those of past performance-based sorting.

Table 11: Mean Group Bonus per Participant

	Willingness-Based	Performance-Based	*p-value*
Decision 1 (STEM&Math)	7.46 (0.17)	7.49 (0.17)	*0.8085*
Decision 2 (STEM&Math)	6.47 (0.19)	6.38 (0.19)	*0.2924*
Decision 3 (STEM&Math)	6.09 (0.18)	6.08 (0.18)	*0.7723*
Decision 1 (non-STEM&Verbal)	6.41 (0.16)	6.66 (0.13)	*0.0192*
Decision 2 (non-STEM&Verbal)	6.15 (0.16)	6.26 (0.15)	*0.5018*
Decision 3 (non-STEM&Verbal)	6.12 (0.16)	6.17 (0.16)	*0.6854*

Notes: Group bonus means for each decision. Standard deviations are in parentheses. P-Values are based on non-parametric Wilcoxon matched-pairs signed-rank test.

It is noteworthy that willingness-based sorting yields higher earnings in math task among STEM students, although the difference from performance-based sorting is not statistically significant. This finding highlights the possibility that performance on the easy task is not necessarily a good signal for the hard task and additionally that participants are good at identifying their future performance on the harder task. Additionally, a gap in promotion between STEM men and women is found in Decision 2, consistent with the previous findings of this study. In this decision, only 37% of the promoted participants are female, and the difference between promoted men and promoted women is significant at the 1% level (Wilcoxon rank-sum (Mann-Whitney) test, $z = 2.988$, $p = 0.0028$).

4.6 Closing the Gap

Presenting STEM Students with a Difficult Math Task First Effectively Closes the Gender Gap in Their Preference for Easier Tasks

Following the results of our first experiment, we designed and pre-registered a new experiment, HighDifficulty, intending to close the gender gap in promotion applications. Based on the previous literature on gender differences in ambiguity aversion, we hypothesized that the ambiguity related to the content of the promotional task might be the driving factor behind the gender difference. Our second experiment was run at the WZB-TU lab of the Technical University of Berlin only hiring STEM students. In this experiment, we only concentrated on the male task where we found the main difference in our first experiment.

When participants have the chance to experience the hard task first, the gender gap is no longer present (See Table 12). Therefore, we conclude that task ambiguity is the driver behind the gender difference initially found in Baseline among STEM students in the male domain. We additionally show that our new experiment closes the promotion gap in the second decision that was identified in Baseline. In HighDifficulty, 49% of the female participants are promoted in the second decision and the difference between men and women is no longer significant (Wilcoxon rank-sum (Mann-Whitney) test, $z = 0.748$, $p = 0.4545$). This increase in female promotions does not trigger any new inefficiency in terms of total group bonus that did not exist in Baseline (Wilcoxon signed-rank test, $z = 1.001$, $p = 0.3169$).

Table 12: The Gender Gap in STEM Students in HighDifficulty

	(1) Baseline	(2) HighDifficulty
Female	0.1848	0.0117
	(0.1775)	(0.1715)
Easy-Hard	0.3500**	0.5647***
	(0.1219)	(0.1262)
Female x *Easy-Hard*	0.3623*	0.0654
	(0.1801)	(0.1969)
Performance	-0.0890	-0.1607*
	(0.0768)	(0.0669)
Controls	YES	YES
Observations	306	316

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: GLS random-effect model with robust standard errors. The dependent variable is the willingness to apply, pooling the second and third decisions. Controls include age, education, employment status, German origins, session fixed effects and task order variables.

5 Conclusion

Our study shows that STEM women refrain from applying for a promotion when easier alternatives are available, which significantly affects the final sorting, making women lose promotion opportunities. We successfully demonstrate that decreasing task difficulty-related ambiguity manages to close this gender gap and increases the sorting efficiency. It emphasizes two important factors. The study field should be considered when analyzing and addressing gender differences in economic decision making. Additionally, the lack of clarity about the difficulty of the task required from promoted candidates appears to be a key factor driving female promotion application behavior. Hence, in our study, in line with other previous findings, providing more information closes the gender gap.

Following our findings, we propose employers draw clear descriptions of the content of the promotional positions for their employees. Some ideas could concern trial periods before promotion applications, increased exposure to managerial tasks and transparency on expectations. Our intervention closes the gap by moving female willingness to apply higher and male lower. Therefore, it also shows the effect of such policies in equalizing the talent pool without making it smaller.

It is important to emphasize, however, that a limitation of our study is its abstract design, which may not fully represent a real-life promotion application setting. However, we know that hierarchical job segregation begins with the first promotion application and increases over the course of an individual's career. Therefore, we do not consider our subject pool to be a limitation. Moreover, we take advantage of the abstract setup to detect the self-limiting behavior that cannot be observed independently of the many confounding factors that occur in real-life settings. Thus, we further demonstrate the ever-present importance of laboratory experiments in policy-making, as any real-world data would have difficulty separating preferences from self-limiting behavior. Moreover, thanks to the lab setup, we also provide panel data where we can observe the behavior of STEM and non-STEM students in both male-dominated and female-dominated tasks, whereas it would not be easy to find setups where engineers (as an example of a male-dominated domain) apply for human resources (as an example of a female-dominated domain) in real life.

Our study further sheds light on the importance of considering background factors in gender studies and therefore fills a crucial gap that can aid in further scrutiny of gender differences. Prior experiments that did not control for the participant's field of study may have either overestimated or underestimated gender differences, depending on the distribution of the participant pool. Our findings indicate that the primary gender differences in confidence are driven by subjects in the STEM field. This finding contradicts the positive selection arguments and adds to the discussion started by Saygin and Atwater (2021) underlining that this preassumed selection might not hold.

Another limitation of our findings is that we can only provide correlational evidence of the effect of STEM fields. We find that the differences only appear between men and women in STEM fields, and whether this difference is due to special selection into these fields from the beginning or whether it is the result of participation in these male-dominated studies. Therefore, we see great value in further research, as these different possible explanations call for different policies to close the gaps.

References

Abraham, L., J. Hallermeier, and A. Stein (2023). Words matter: Experimental evidence from job applications.

Barbulescu, R. and M. Bidwell (2013). Do women choose different jobs from men? mechanisms of application segregation in the market for managerial workers. *Organization Science 24*(3), 737–756.

Barth, E., S. P. Kerr, and C. Olivetti (2021). The dynamics of gender earnings differentials: Evidence from establishment data. *European Economic Review 134*, 103713.

Bettio, F., A. Verashchagina, and F. Camilleri-Cassar (2009). Gender segregation in the labour market: Root causes, implications and policy responses in the eu.

Blau, F. D. and J. DeVaro (2007). New evidence on gender differences in promotion rates: An empirical analysis of a sample of new hires. *Industrial Relations: A Journal of Economy and Society 46*(3), 511–550.

Blau, F. D. and L. M. Kahn (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature 55*(3), 789–865.

Bohnet, I., A. Van Geen, and M. Bazerman (2016). When performance trumps gender bias: Joint vs. separate evaluation. *Management Science 62*(5), 1225–1234.

Bosquet, C., P.-P. Combes, and C. García-Peñalosa (2019). Gender and promotions: Evidence from academic economists in france. *The Scandinavian Journal of Economics 121*(3), 1020–1053.

Brands, R. A. and I. Fernandez-Mateo (2017). Leaning out: How negative recruitment experiences shape women's decisions to compete for executive roles. *Administrative Science Quarterly 62*(3), 405–442.

Brenøe, A. A., L. Heursen, E. Ranehill, and R. A. Weber (2022). Continuous gender identity and economics. In *AEA Papers and Proceedings*, Volume 112, pp. 573–77.

Buser, T., M. Niederle, and H. Oosterbeek (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics 129*(3), 1409–1447.

Card, D. and A. B. Krueger (1994). Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania. *American Economic Review 84*(4), 772–793.

Charness, G., U. Gneezy, and V. Rasocha (2021). Experimental methods: Eliciting beliefs. *Journal of Economic Behavior & Organization 189*, 234–256.

Chrousos, G. P. and A.-F. A. Mentis (2020). Imposter syndrome threatens diversity. *Science 367*(6479), 749–750.

Clance, P. R. and S. A. Imes (1978). The imposter phenomenon in high achieving women: Dynamics and therapeutic intervention. *Psychotherapy: Theory, Research & Practice 15*(3), 241.

Clark Blickenstaff, J. (2005). Women and science careers: leaky pipeline or gender filter? *Gender and Education 17*(4), 369–386.

Coffman, K. B. (2014). Evidence on self-stereotyping and the contribution of ideas. *The Quarterly Journal of Economics 129*(4), 1625–1660.

Coffman, K. B., M. R. Collis, and L. Kulkarni (2022). Whether to apply.

Croson, R. and U. Gneezy (2009). Gender differences in preferences. *Journal of Economic Literature 47*(2), 448–474.

Dickerson, A. and M. A. Taylor (2000). Self-limiting behavior in women: Self-esteem and self-efficacy as predictors. *Group & Organization Management 25*(2), 191–210.

Dreber, A., E. Von Essen, and E. Ranehill (2014). Gender and competition in adolescence: task matters. *Experimental Economics 17*, 154–172.

Eckel, C. C. and P. J. Grossman (2008). Men, women and risk aversion: Experimental evidence. *Handbook of Experimental Economics Results 1*, 1061–1073.

Encinas-Martín, M. and M. Cherian (2023). *Gender, Education and Skills.*

Exley, C. L. and J. B. Kessler (2022). The gender gap in self-promotion. *The Quarterly Journal of Economics 137*(3), 1345–1381.

Fallucchi, F., D. Nosenzo, and E. Reuben (2020). Measuring preferences for competition with experimentally-validated survey questions. *Journal of Economic Behavior & Organization 178*, 402–423.

Flory, J. A., A. Leibbrandt, and J. A. List (2015). Do competitive workplaces deter female workers? a large-scale natural field experiment on job entry decisions. *The Review of Economic Studies 82*(1), 122–155.

Gee, L. K. (2019). The more you know: Information effects on job application rates in a large field experiment. *Management Science 65*(5), 2077–2094.

Gneezy, U., M. Niederle, and A. Rustichini (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics 118*(3), 1049–1074.

Goebel, J., M. M. Grabka, S. Liebig, M. Kroh, D. Richter, C. Schröder, and J. Schupp (2019). The german socio-economic panel (soep). *Jahrbücher für Nationalökonomie und Statistik 239*(2), 345–360.

Mohr, T. S. (2014). Why women don't apply for jobs unless they're 100% qualified. *Harvard Business Review 25*, 40–45.

Moore, D. A. and P. J. Healy (2008). The trouble with overconfidence. *Psychological Review 115*(2), 502.

Samek, A. (2019). Gender differences in job entry decisions: A university-wide field experiment. *Management Science 65*(7), 3272–3281.

Sarsons, H. (2017). Recognition for group work: Gender differences in academia. *American Economic Review 107*(5), 141–145.

Saygin, P. O. and A. Atwater (2021). Gender differences in leaving questions blank on high-stakes standardized tests. *Economics of Education Review 84*, 102162.

Whitman, M. V. and K. K. Shanine (2012). Revisiting the impostor phenomenon: How individuals cope with feelings of being in over their heads. In *The role of the economic crisis on occupational stress and well being*, Volume 10, pp. 177–212. Emerald Group Publishing Limited.

A Appendix

A.1 Supplementary Analyses

A.1.1 Past Performance Expectations in Baseline

Table 13: Past Performance Expectations in BASELINE Pooled

	Math		Verbal	
	(Abs)	(Rel)	(Abs)	(Rel)
Male	0.493***	0.444**	0.223	0.242
	(0.130)	(0.151)	(0.132)	(0.153)
Performance	0.248***	0.307***	0.225***	0.157**
	(0.0595)	(0.0699)	(0.0515)	(0.0547)
Controls	YES	YES	YES	YES
Obs	303	303	303	303
R^2	0.285	0.295	0.208	0.151

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS model with robust standard errors. The dependent variables are elicited past performance expectations about absolute and relative performances. Controls include age, education, employment status, German origins, session fixed effects and task order variables.

Table 14: Past Performance Expectations in BASELINE by Gender

	Math				Verbal			
	Men		Women		Men		Women	
	(Abs)	(Rel)	(Abs)	(Rel)	(Abs)	(Rel)	(Abs)	(Rel)
STEM	0.228	0.614*	0.429	0.543	0.453*	0.687**	0.00259	0.337
	(0.234)	(0.277)	(0.271)	(0.284)	(0.219)	(0.259)	(0.243)	(0.251)
Performance	0.172*	0.431***	0.267**	0.132	0.214**	0.0978	0.201*	0.149
	(0.0784)	(0.102)	(0.0883)	(0.102)	(0.0765)	(0.0774)	(0.0809)	(0.0789)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Obs	150	150	153	153	150	150	153	153
R^2	0.359	0.408	0.328	0.357	0.308	0.289	0.315	0.301

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS model with robust standard errors. The dependent variables are elicited past performance expectations about absolute and relative performances. Controls include age, education, employment status, German origins, session fixed effects and task order variables.

Table 15: Past Performance Expectations in Baseline Comparing STEM Men to Non-STEM Women

	Math		Verbal	
	(Abs)	(Rel)	(Abs)	(Rel)
Male	0.949***	1.293***	0.273	0.586*
	(0.229)	(0.269)	(0.212)	(0.245)
Performance	0.240**	0.108	0.306***	0.211*
	(0.0863)	(0.109)	(0.0743)	(0.0831)
Controls	YES	YES	YES	YES
Obs	160	160	160	160
R^2	0.409	0.407	0.277	0.227

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS model with robust standard errors. The dependent variables are elicited past performance expectations about absolute and relative performances. Controls include age, education, employment status, German origins, session fixed effects and task order variables.

A.1.2 Future Performance Expectations in a Familiar (Easy) Task in Baseline

Table 16: Future Performance Expectations in a Familiar (Easy) Task in Baseline by Field of Study

	Math				Verbal			
	STEM		Non-STEM		STEM		Non-STEM	
	(Abs)	(Rel)	(Abs)	(Rel)	(Abs)	(Rel)	(Abs)	(Rel)
Male	0.480**	0.596**	0.526*	0.389	0.151	0.124	0.0876	0.328
	(0.155)	(0.206)	(0.223)	(0.245)	(0.183)	(0.204)	(0.218)	(0.262)
Performance	0.207**	0.308**	0.334***	0.295**	0.145*	0.107	0.305***	0.224**
	(0.0738)	(0.0971)	(0.0938)	(0.110)	(0.0708)	(0.0729)	(0.0661)	(0.0852)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Obs	153	153	150	150	153	153	150	150
R^2	0.456	0.390	0.487	0.390	0.324	0.299	0.366	0.278

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS model with robust standard errors. The dependent variables are elicited future performance expectations in an easy task about absolute and relative performances. Controls include age, education, employment status, German origins, session fixed effects and task order variables.

Table 17: Future Performance Expectations in a Familiar (Easy) Task in BASELINE by Gender

	Math				Verbal			
	Men		Women		Men		Women	
	(Abs)	(Rel)	(Abs)	(Rel)	(Abs)	(Rel)	(Abs)	(Rel)
STEM	0.277	0.818**	0.452	0.570	0.431	0.645*	0.115	0.365
	(0.220)	(0.260)	(0.263)	(0.288)	(0.221)	(0.273)	(0.217)	(0.250)
Performance	0.263**	0.471***	0.321***	0.178	0.253***	0.193*	0.188**	0.0960
	(0.0862)	(0.106)	(0.0821)	(0.107)	(0.0646)	(0.0793)	(0.0704)	(0.0700)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Obs	150	150	153	153	150	150	153	153
R^2	0.413	0.422	0.406	0.383	0.364	0.311	0.353	0.292

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS model with robust standard errors. The dependent variables are elicited future performance expectations in a hard task about absolute and relative performances. Controls include age, education, employment status, German origins, session fixed effects and task order variables.

A.1.3 Future Performance Expectations in a Hard Task in Baseline

Table 18: Future Performance Expectations in a Hard Task in BASELINE by Field of Study

	Math				Verbal			
	STEM		Non-STEM		STEM		Non-STEM	
	(Abs)	(Rel)	(Abs)	(Rel)	(Abs)	(Rel)	(Abs)	(Rel)
Male	0.489**	0.597**	0.322	0.224	0.373	0.350	0.0495	0.251
	(0.182)	(0.205)	(0.234)	(0.259)	(0.215)	(0.225)	(0.229)	(0.275)
Performance	0.171*	0.375***	0.132	0.145	0.0872	0.165	0.207**	0.203*
	(0.0821)	(0.0966)	(0.0932)	(0.112)	(0.0870)	(0.0854)	(0.0686)	(0.100)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Obs	153	153	150	150	153	153	150	150
R^2	0.346	0.423	0.332	0.341	0.268	0.310	0.327	0.292

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS model with robust standard errors. The dependent variables are elicited future performance expectations in a hard task about absolute and relative performances. Controls include age, education, employment status, German origins, session fixed effects and task order variables.

Table 19: Future Performance Expectations in a Hard Task in BASELINE by Gender

	Math				Verbal			
	Men		Women		Men		Women	
	(Abs)	(Rel)	(Abs)	(Rel)	(Abs)	(Rel)	(Abs)	(Rel)
STEM	0.449	1.071***	0.205	0.605	0.491	0.666*	-0.189	0.343
	(0.256)	(0.288)	(0.265)	(0.317)	(0.308)	(0.324)	(0.240)	(0.269)
Performance	0.127	0.406***	0.181*	0.126	0.201*	0.329***	0.126	0.0893
	(0.0871)	(0.104)	(0.0866)	(0.115)	(0.0819)	(0.0940)	(0.0731)	(0.0803)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Obs	150	150	153	153	150	150	153	153
R^2	0.385	0.501	0.344	0.317	0.301	0.339	0.267	0.319

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS model with robust standard errors. The dependent variables are elicited future performance expectations in a hard task about absolute and relative performances. Controls include age, education, employment status, German origins, session fixed effects and task order variables.

Table 20: Future Performance Expectations in a Hard Task in BASELINE Comparing Non-STEM Men to STEM Women

	Math		Verbal	
	(Abs)	(Rel)	(Abs)	(Rel)
Male	0.196	-0.145	0.280	-0.00869
	(0.266)	(0.333)	(0.331)	(0.385)
Performance	0.185*	0.448***	0.101	0.150
	(0.0828)	(0.0998)	(0.0767)	(0.0937)
Controls	YES	YES	YES	YES
Obs	143	143	143	143
R^2	0.351	0.438	0.268	0.323

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS model with robust standard errors. The dependent variables are elicited future performance expectations in a hard task about absolute and relative performances. Controls include age, education, employment status, German origins, session fixed effects and task order variables.

Table 21: Future Performance Expectations in a Hard Task in BASELINE Comparing STEM Men to Non-STEM Women

	Math		Verbal	
	(Abs)	(Rel)	(Abs)	(Rel)
Male	0.806**	1.656***	0.189	0.661**
	(0.243)	(0.269)	(0.244)	(0.247)
Performance	0.130	0.0883	0.189*	0.210*
	(0.0912)	(0.124)	(0.0751)	(0.0823)
Controls	YES	YES	YES	YES
Obs	160	160	160	160
R^2	0.380	0.457	0.224	0.288

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS model with robust standard errors. The dependent variables are elicited future performance expectations in a hard task about absolute and relative performances. Controls include age, education, employment status, German origins, session fixed effects and task order variables.

A.2 Willingness to Apply for Promotions in Baseline

A.2.1 Pooled Willingness to Apply for Promotions

Table 22: Pooled Willingness to Apply in Baseline

	NS men vs NS women		NS men vs S women		S women vs NS women	
	Math	Verbal	Math	Verbal	Math	Verbal
Male	-0.184	-0.235	-0.0687	-0.138		
	(0.154)	(0.136)	(0.187)	(0.195)		
STEM					-0.215	-0.0454
					(0.173)	(0.170)
Performance	-0.0683	-0.105*	-0.191**	-0.0822	-0.120*	-0.0240
	(0.0581)	(0.0437)	(0.0689)	(0.0483)	(0.0584)	(0.0430)
Controls	YES	YES	YES	YES	YES	YES
Obs	450	450	429	429	480	480

Standard errors in parentheses. NS: Non-STEM, S:STEM.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Random effects model with robust standard errors. The dependent variables are pooled past performance expectations about absolute and relative performances. Controls include age, education, employment status, German origins, session fixed effects and task order variables.

A.2.2 Willingness to Apply for Promotions by Task Difficulty

Table 23: Difference in Differences between STEM Men and non-STEM Women

	Easy-Hard vs *Hard-Hard*		*Easy-Easy* vs *Hard-Hard*	
	Math	Verbal	Math	Verbal
Female	0.559**	-0.166	0.652***	-0.206
	(0.186)	(0.171)	(0.174)	(0.167)
Easy-Hard	0.350**	0.288*		
	(0.122)	(0.116)		
Female x *Easy-Hard*	0.250	0.588***		
	(0.189)	(0.176)		
Easy-Easy			-0.325**	-0.188*
			(0.101)	(0.0921)
Female x *Easy-Easy*			-0.0125	0.113
			(0.158)	(0.130)
Performance	-0.0116	-0.0944*	0.0309	-0.131*
	(0.0574)	(0.0473)	(0.0596)	(0.0526)
Controls	YES	YES	YES	YES
Observations	320	320	320	320

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: GLS random-effect model with robust standard errors. The dependent variable is the willingness to apply pooling the second and third decisions for *Easy-Hard* vs *Hard-Hard*, and the first and third decisions for *Easy-Easy* vs *Hard-Hard*. *Hard Task* stands for the third decision where the novice job also has the hard difficulty questions as the task. Controls include age, correct answers, education, employment status, German origins, university and task order variables.

A.3 Instructions

A.3.1 Experiment 1

Figure 2: Page 1

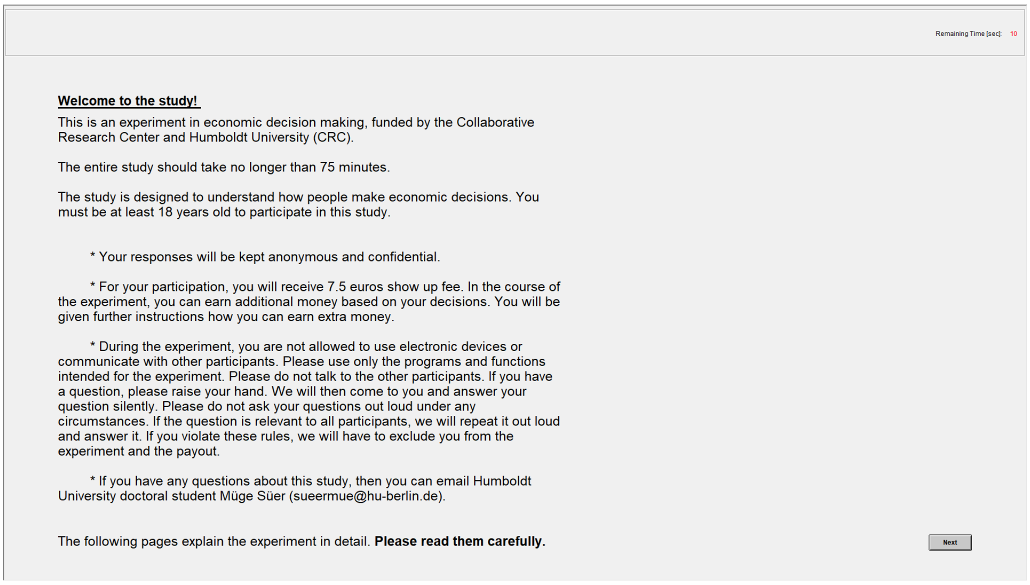


Figure 3: Page 2

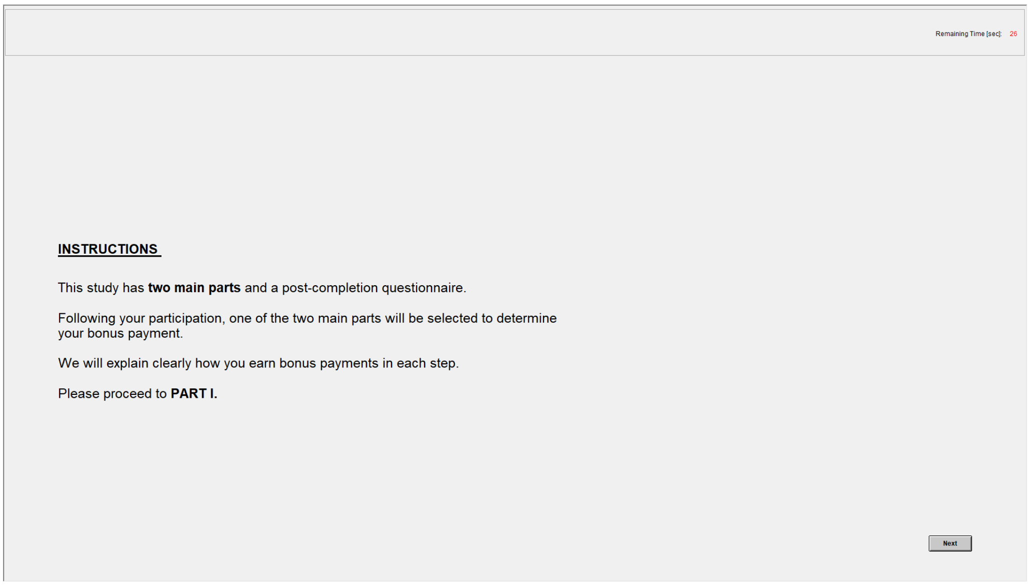


Figure 4: Page 3

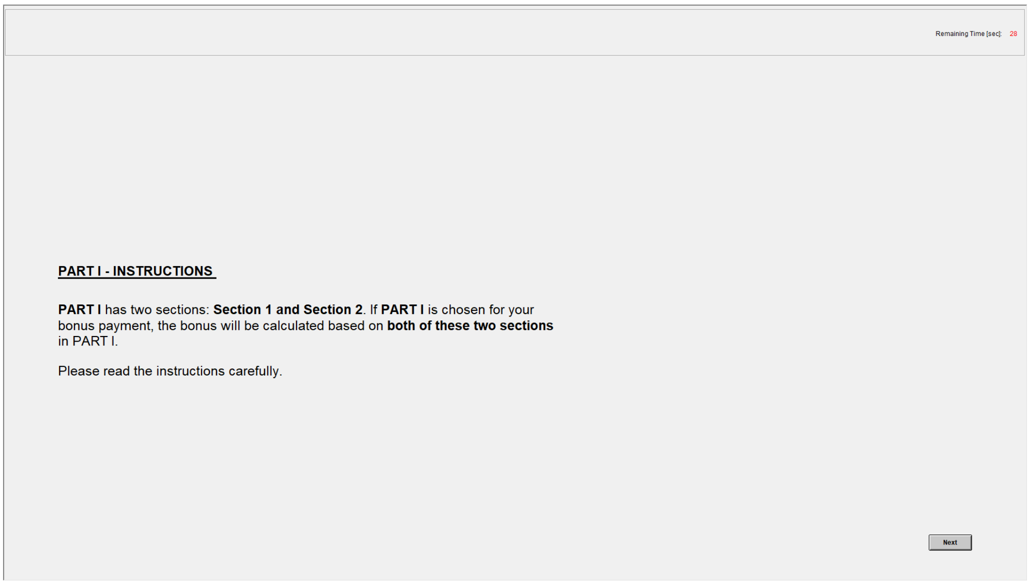


Figure 5: Page 4

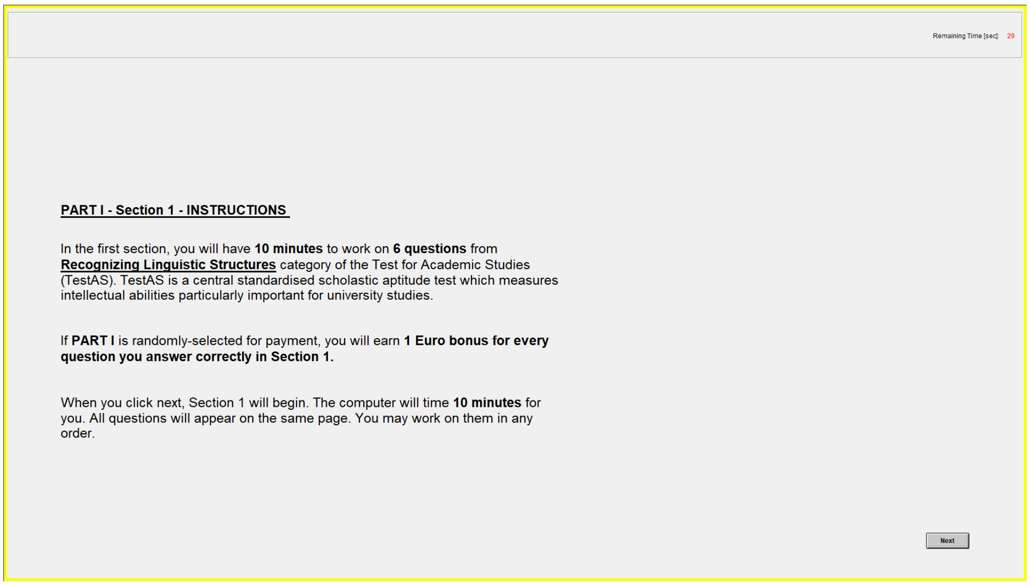


Figure 6: Page 5

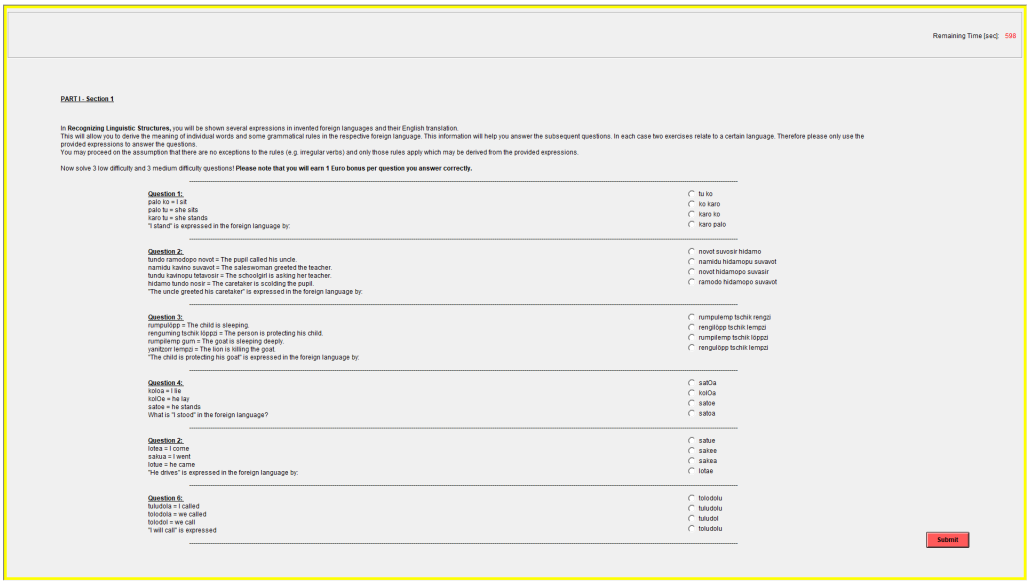


Figure 7: Page 6

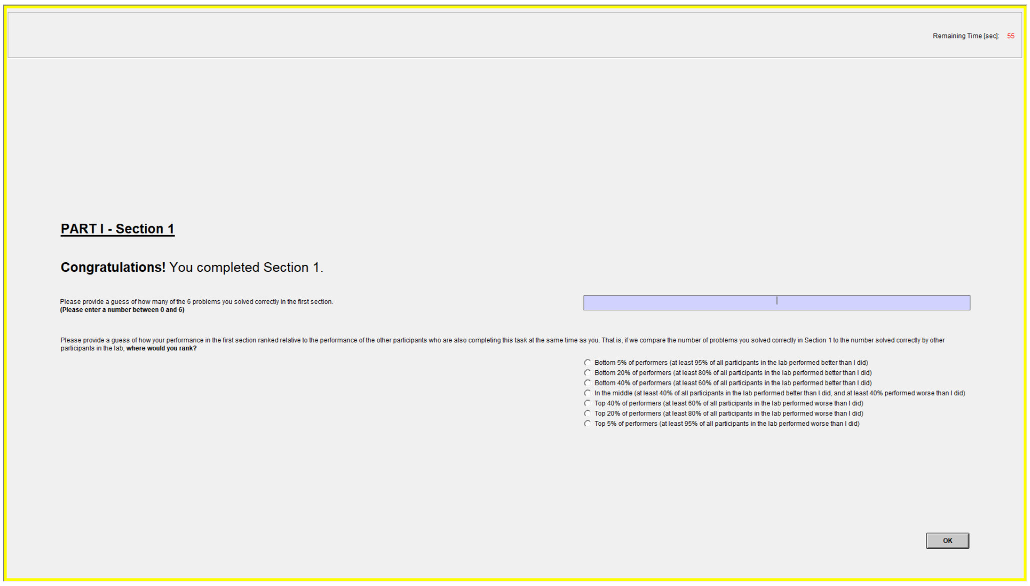


Figure 8: Page 7

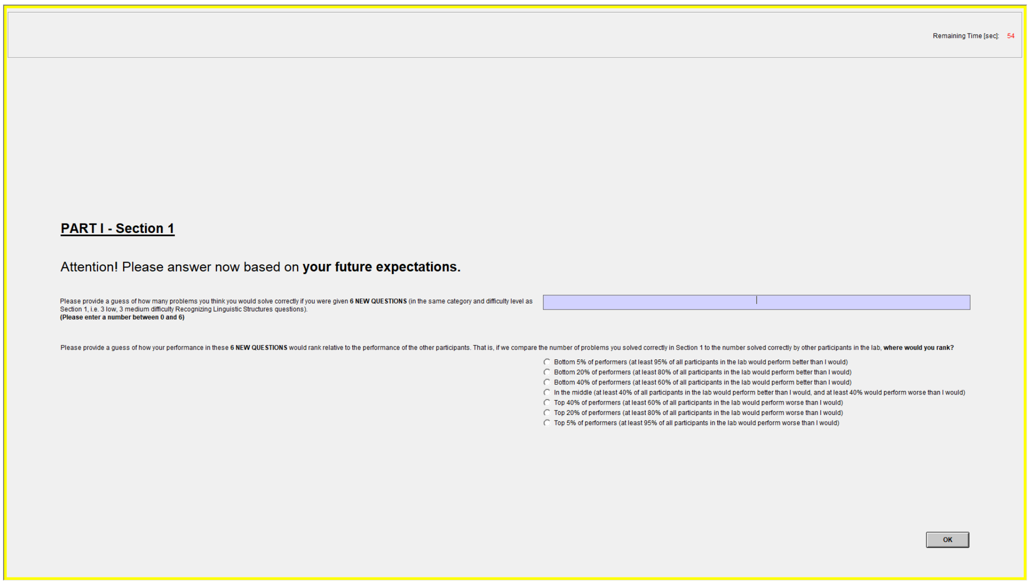


Figure 9: Page 8

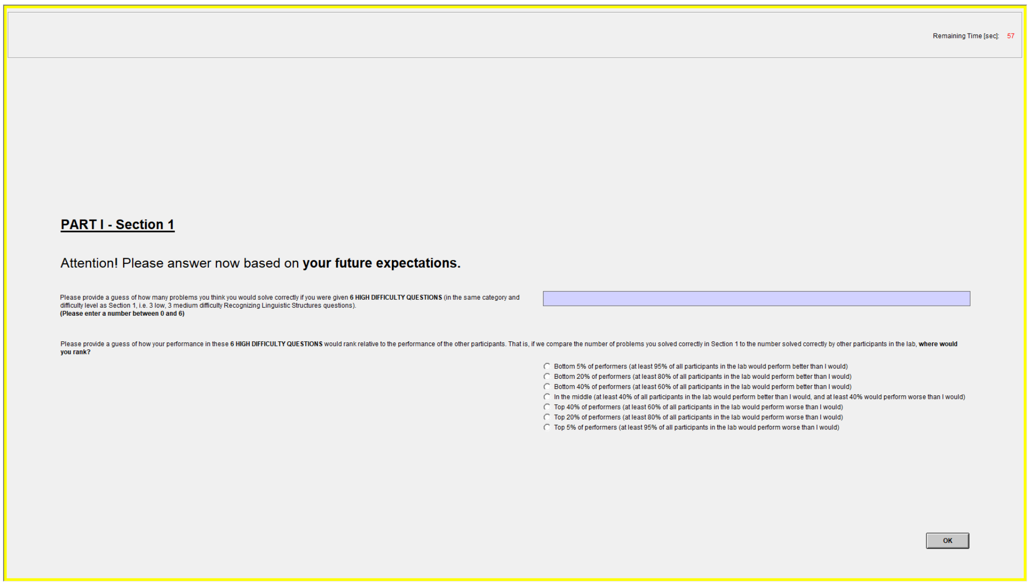
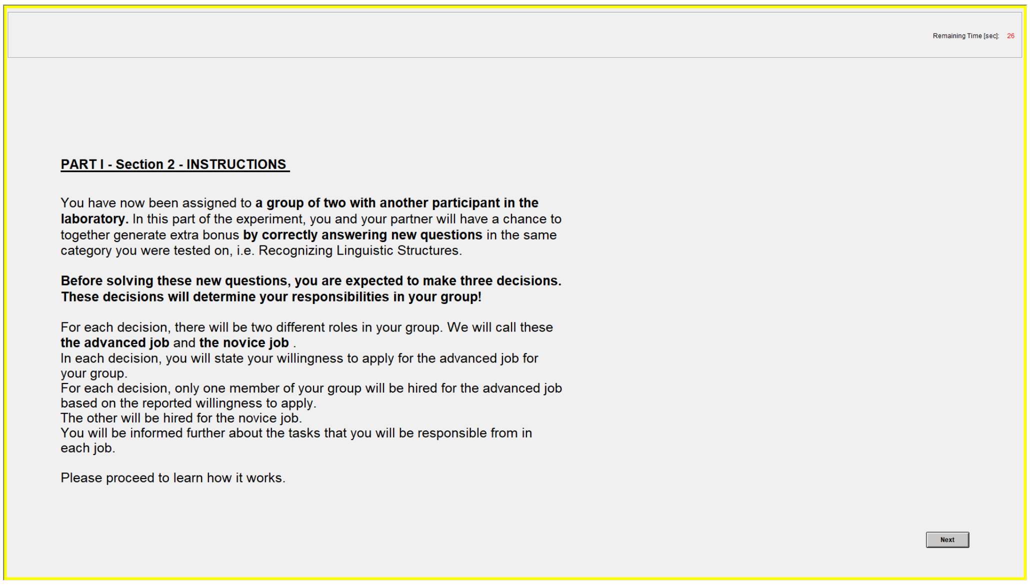


Figure 10: Page 9



Remaining Time [sec]: 26

PART I - Section 2 - INSTRUCTIONS

You have now been assigned to a **group of two with another participant in the laboratory.** In this part of the experiment, you and your partner will have a chance to together generate extra bonus **by correctly answering new questions** in the same category you were tested on, i.e. Recognizing Linguistic Structures.

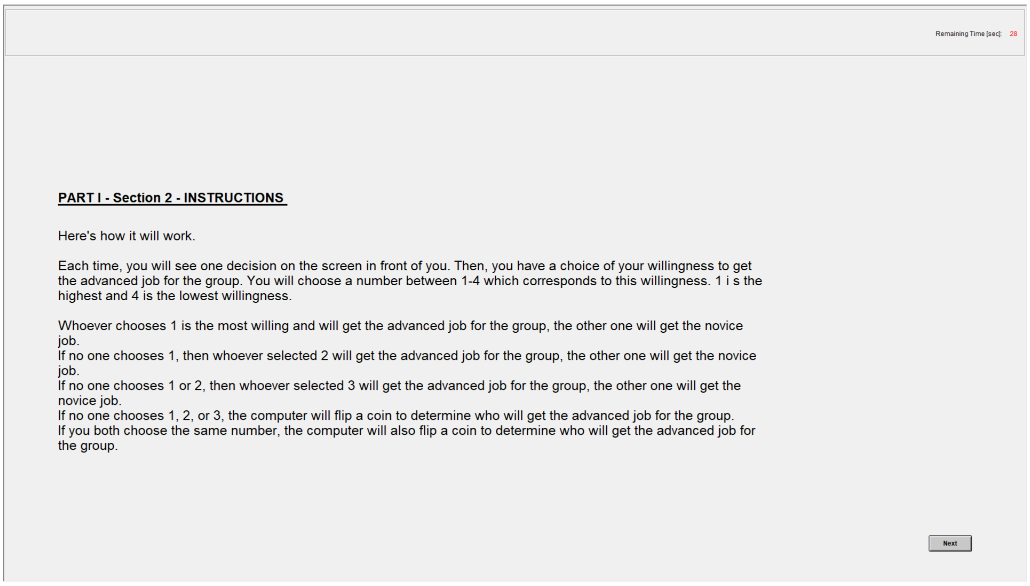
Before solving these new questions, you are expected to make three decisions. These decisions will determine your responsibilities in your group!

For each decision, there will be two different roles in your group. We will call these **the advanced job** and **the novice job** .
In each decision, you will state your willingness to apply for the advanced job for your group.
For each decision, only one member of your group will be hired for the advanced job based on the reported willingness to apply.
The other will be hired for the novice job.
You will be informed further about the tasks that you will be responsible from in each job.

Please proceed to learn how it works.

Next

Figure 11: Page 10



Remaining Time [sec]: 28

PART I - Section 2 - INSTRUCTIONS

Here's how it will work.

Each time, you will see one decision on the screen in front of you. Then, you have a choice of your willingness to get the advanced job for the group. You will choose a number between 1-4 which corresponds to this willingness. 1 i s the highest and 4 is the lowest willingness.

Whoever chooses 1 is the most willing and will get the advanced job for the group, the other one will get the novice job.
If no one chooses 1, then whoever selected 2 will get the advanced job for the group, the other one will get the novice job.
If no one chooses 1 or 2, then whoever selected 3 will get the advanced job for the group, the other one will get the novice job.
If no one chooses 1, 2, or 3, the computer will flip a coin to determine who will get the advanced job for the group.
If you both choose the same number, the computer will also flip a coin to determine who will get the advanced job for the group.

Next

Figure 12: Page 11

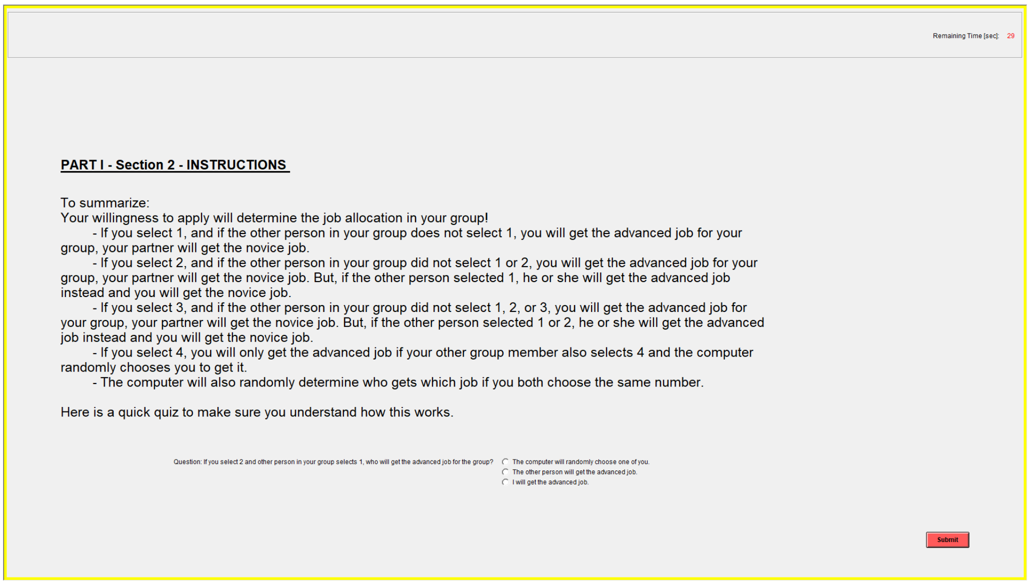


Figure 13: Page 12

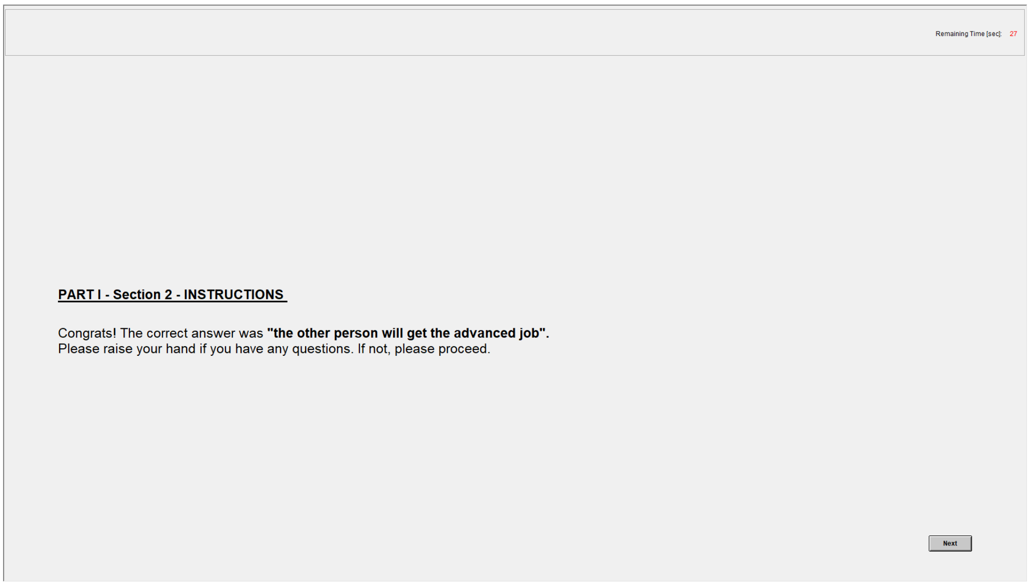


Figure 14: Page 13 (Waitingscreen)

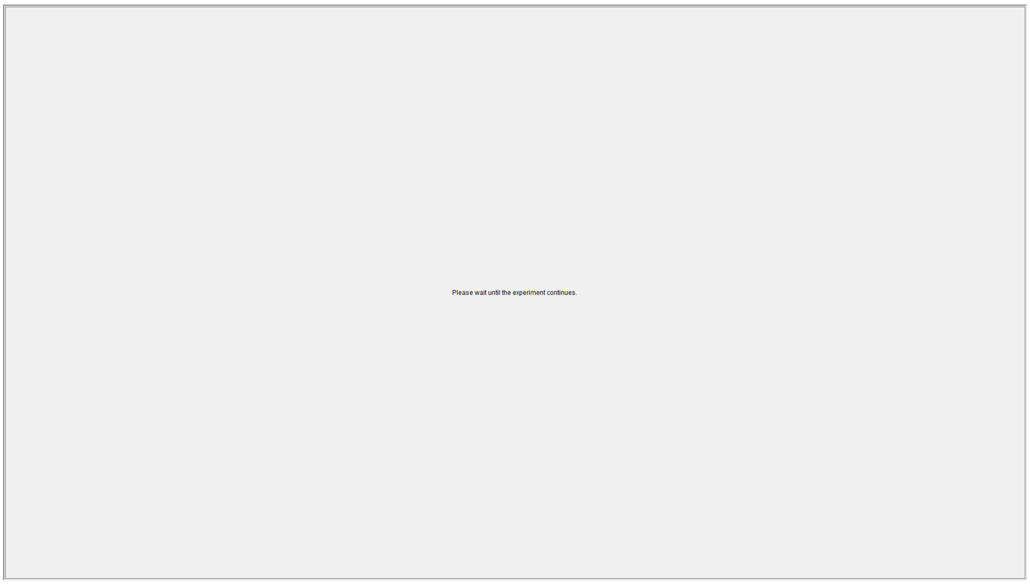


Figure 15: Page 14

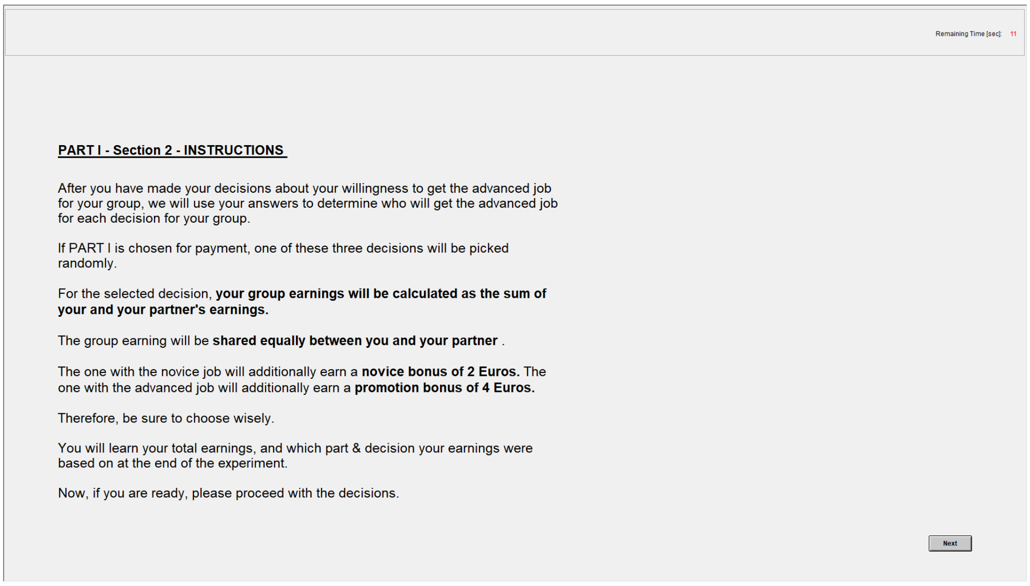


Figure 16: Page 15

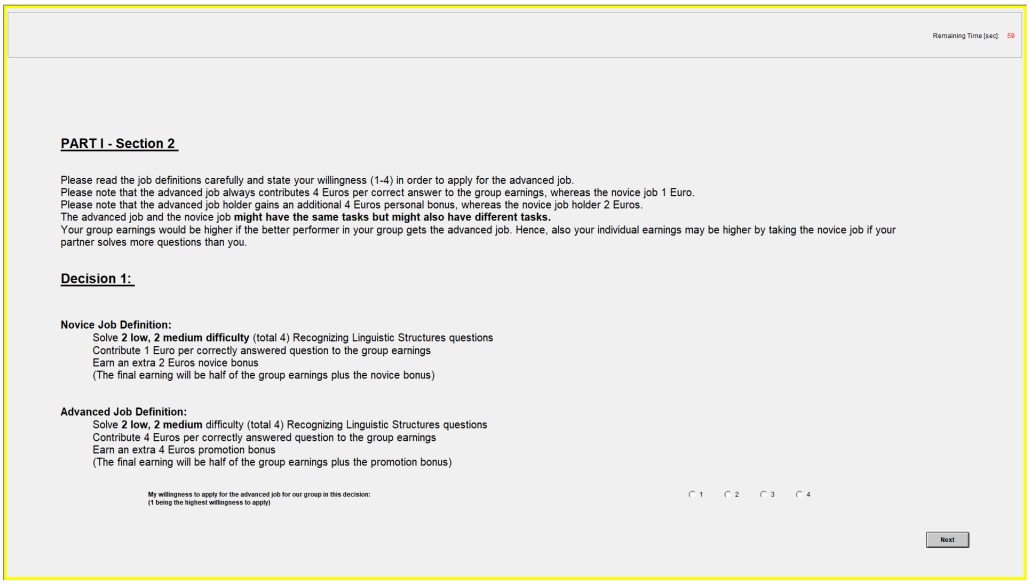


Figure 17: Page 16

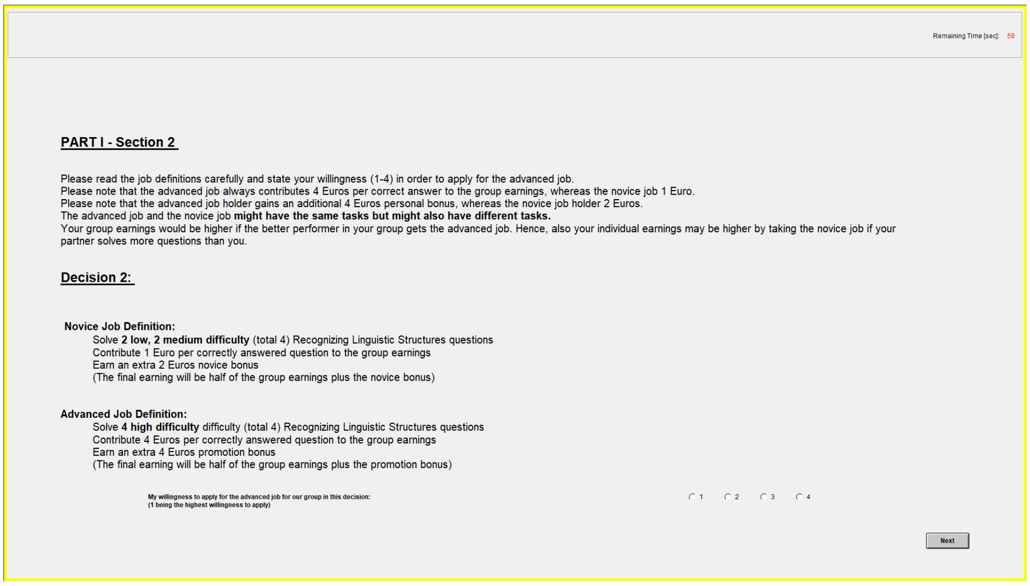


Figure 18: Page 17

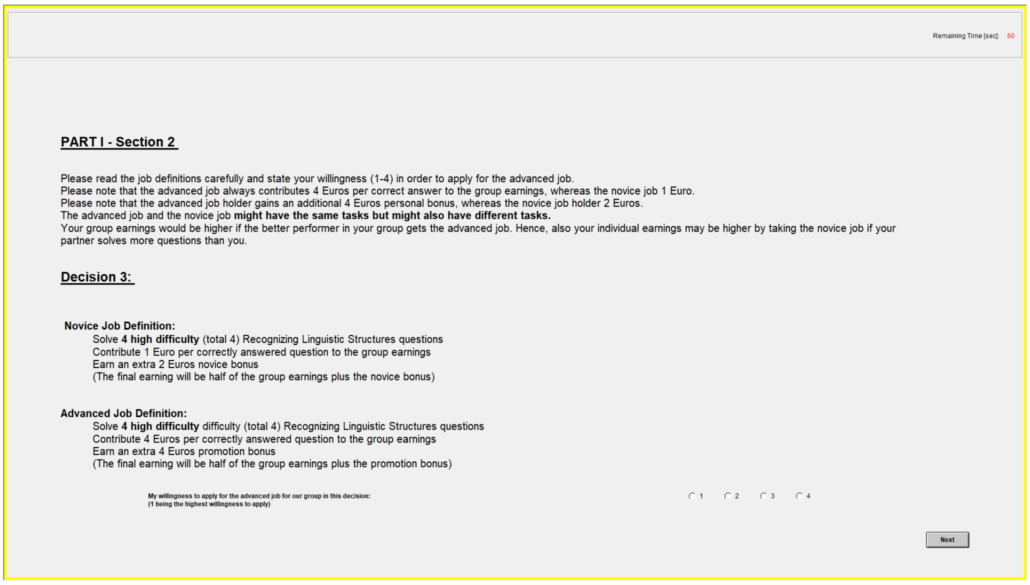


Figure 19: Page 18

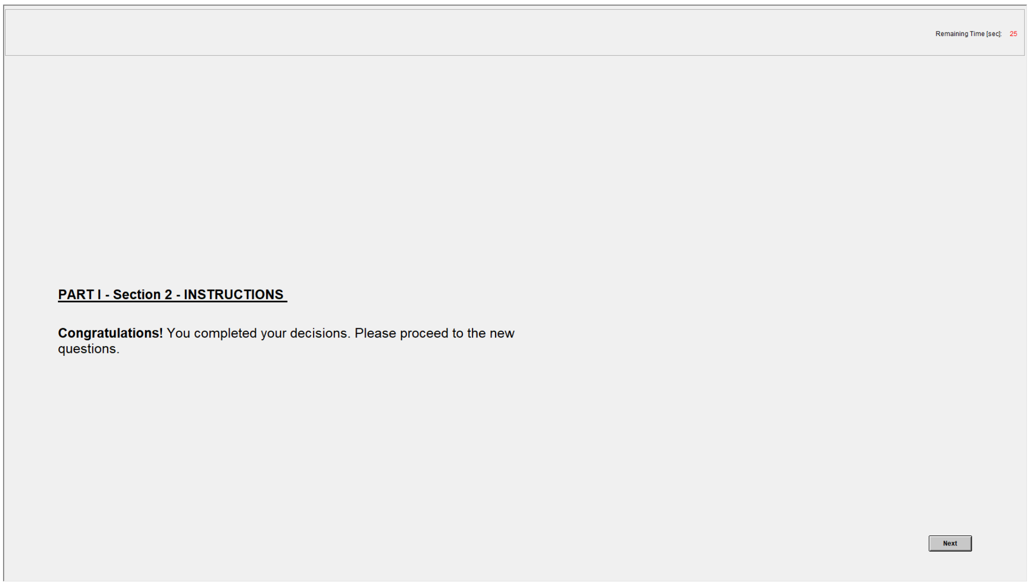


Figure 20: Page 19

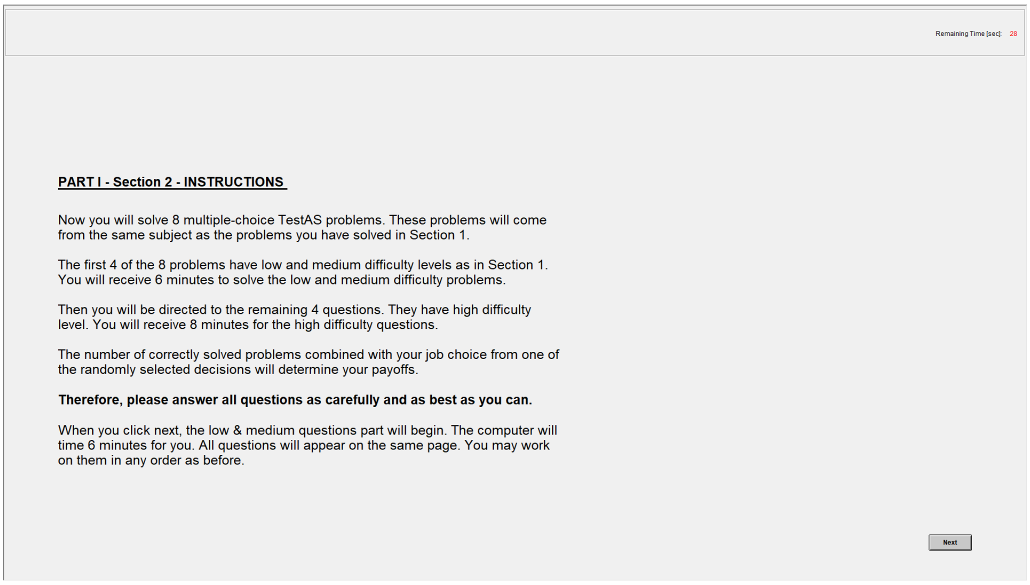


Figure 21: Page 20

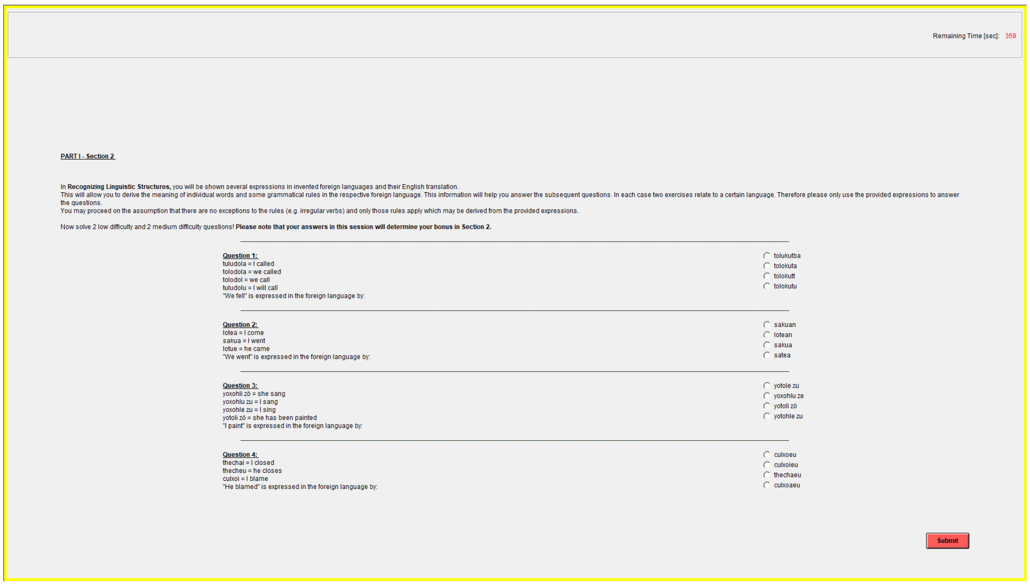


Figure 22: Page 21

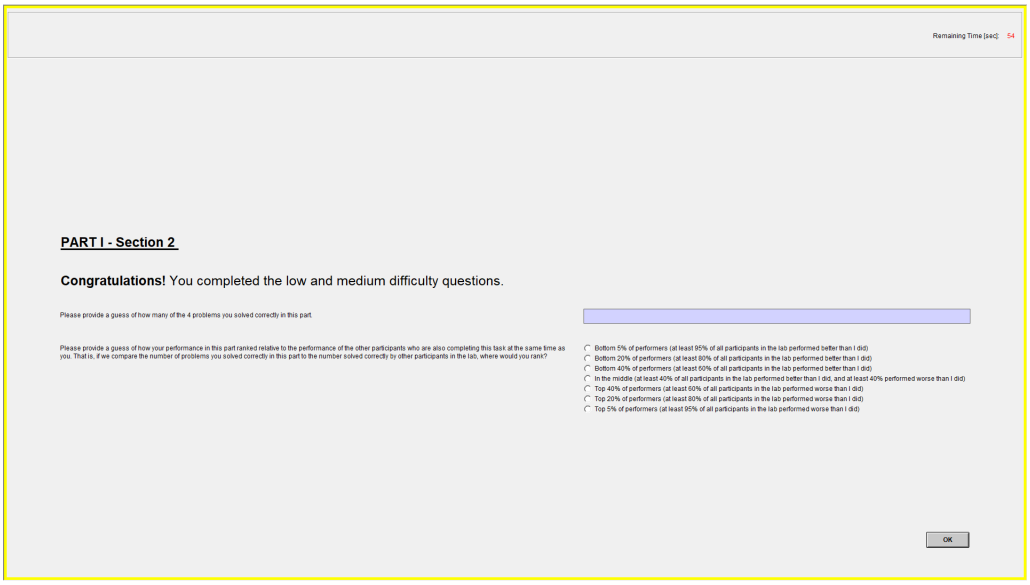


Figure 23: Page 22

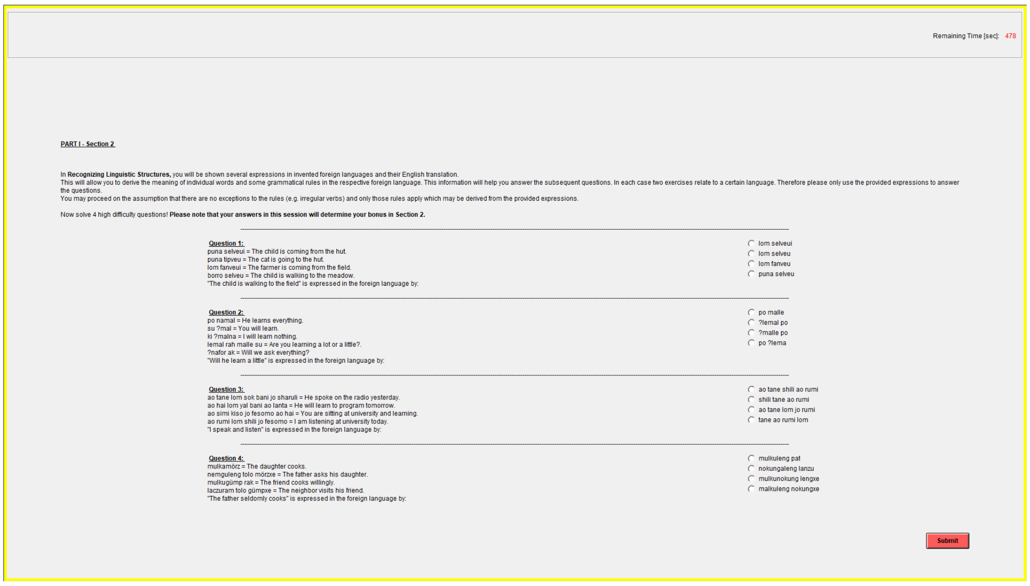


Figure 24: Page 23

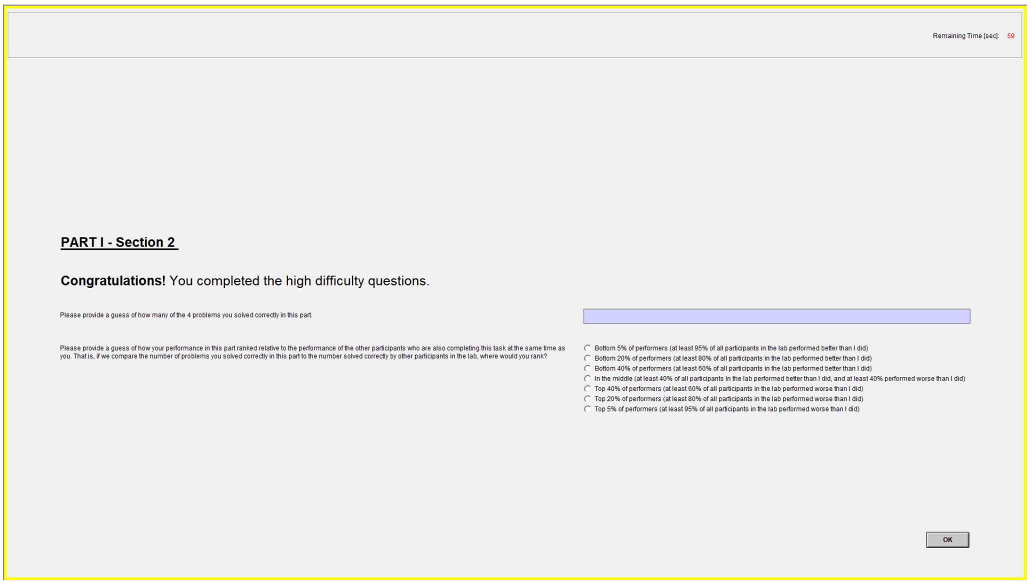


Figure 25: Page 24 (Waitingscreen)

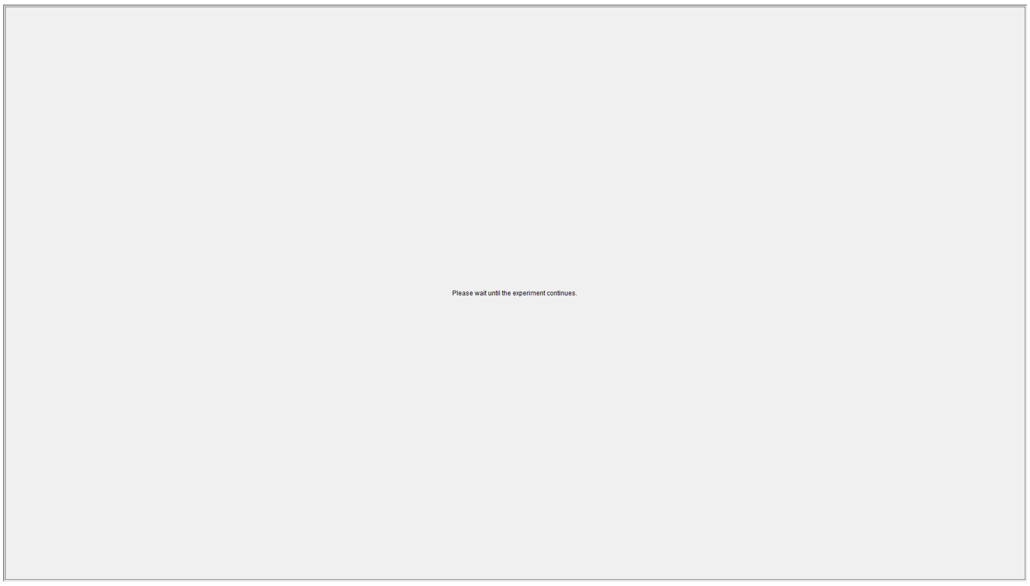


Figure 26: Page 25

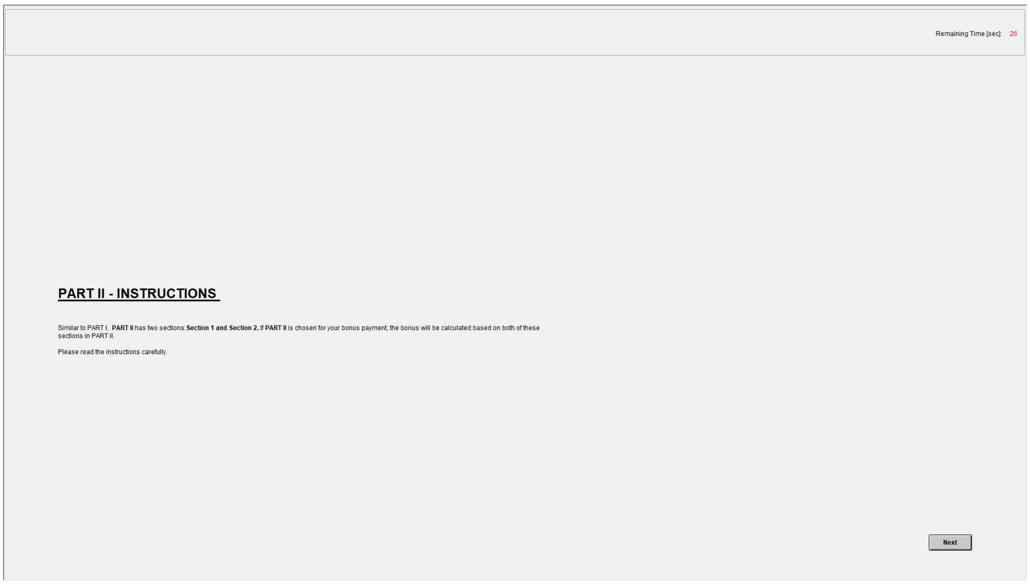


Figure 27: Page 26

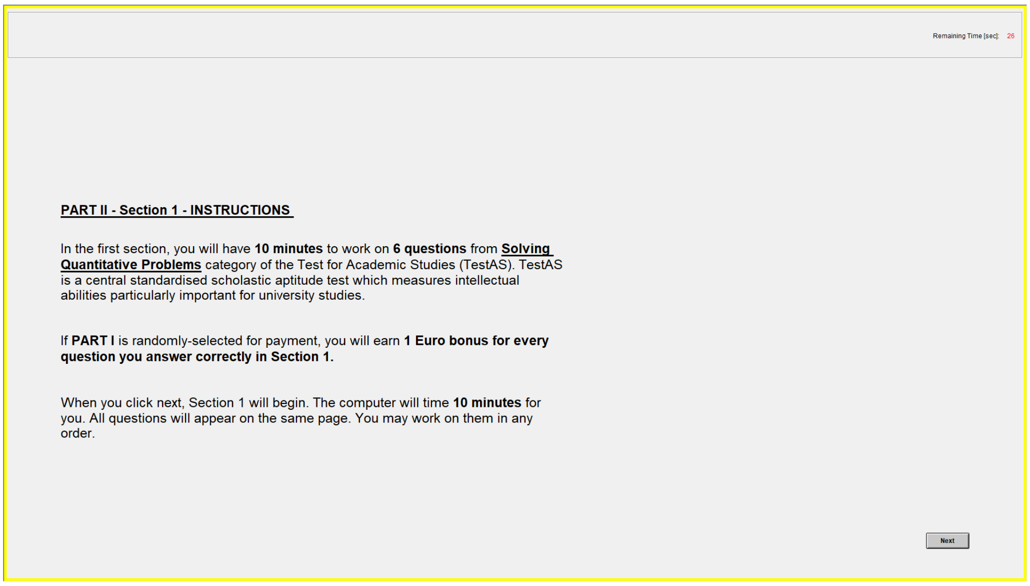


Figure 28: Page 27

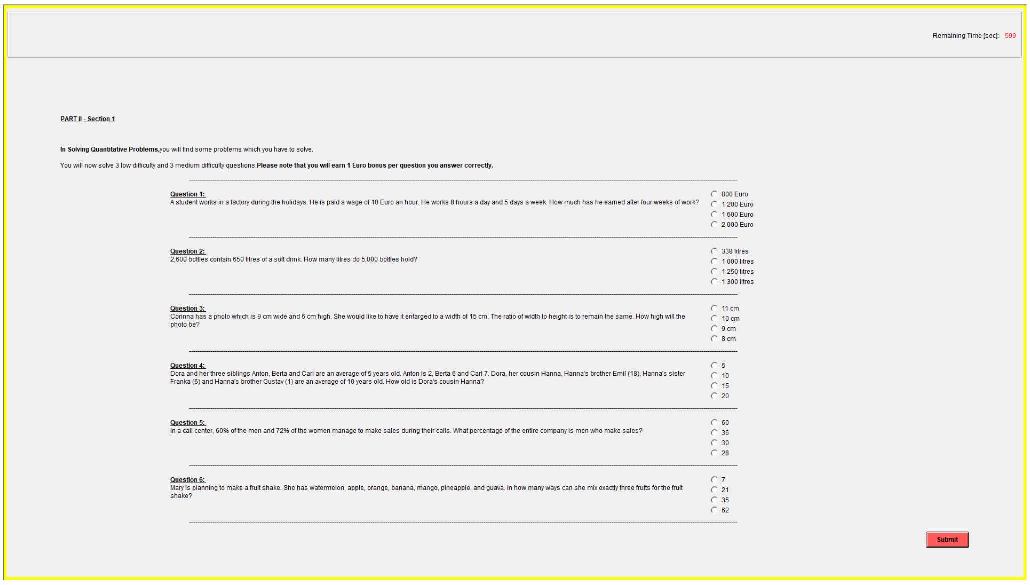


Figure 29: Page 28

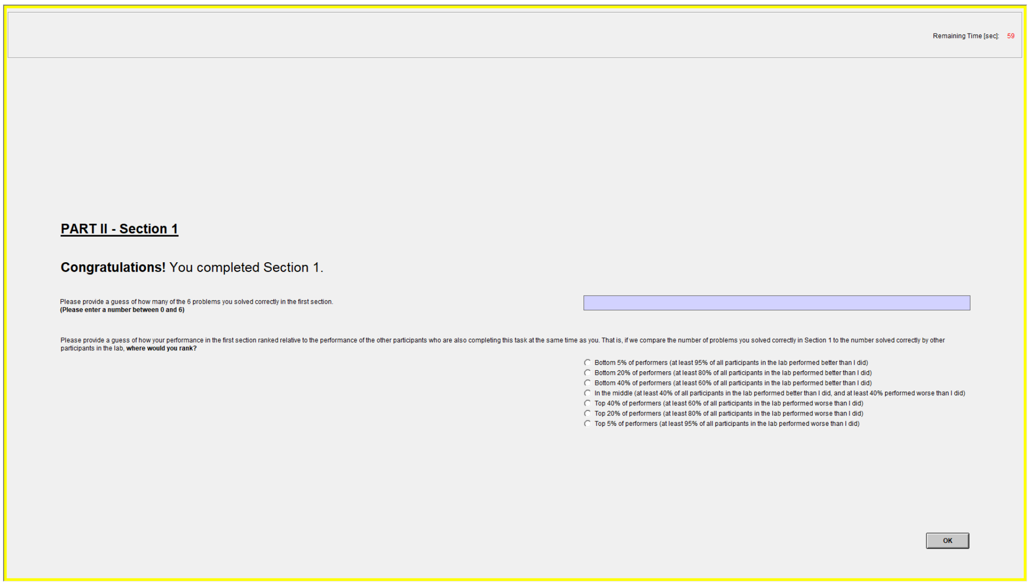


Figure 30: Page 29

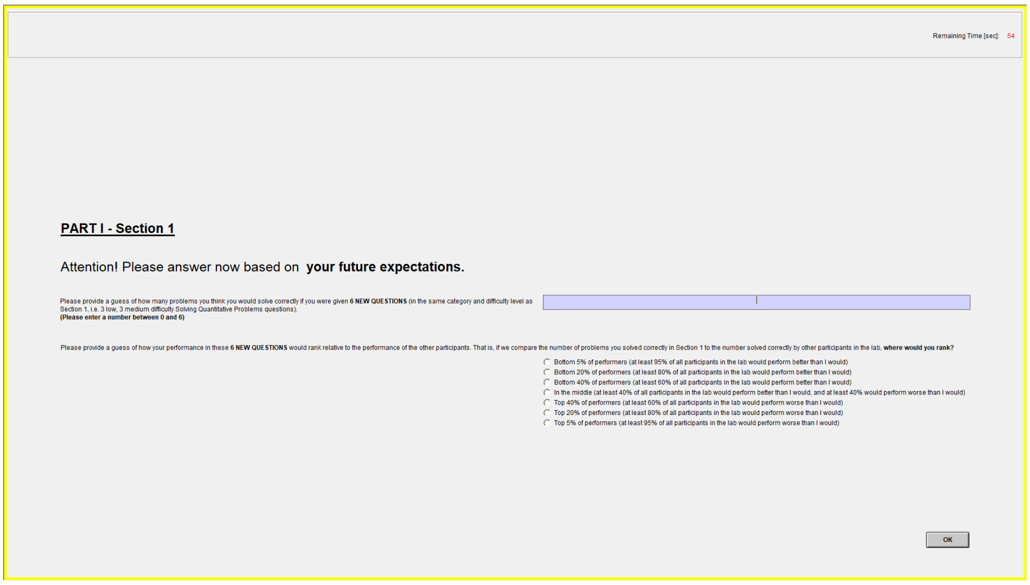


Figure 31: Page 30

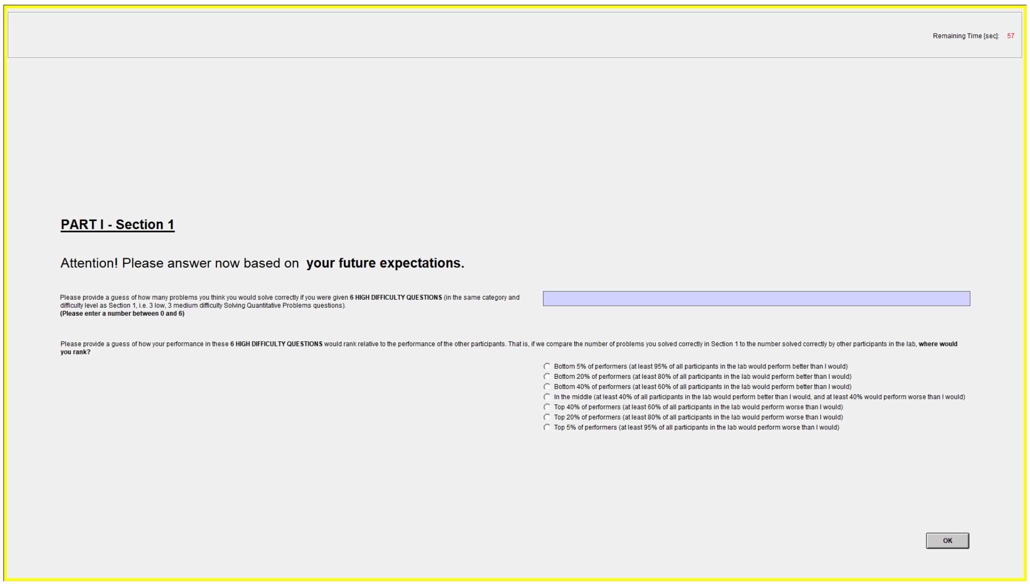


Figure 32: Page 31

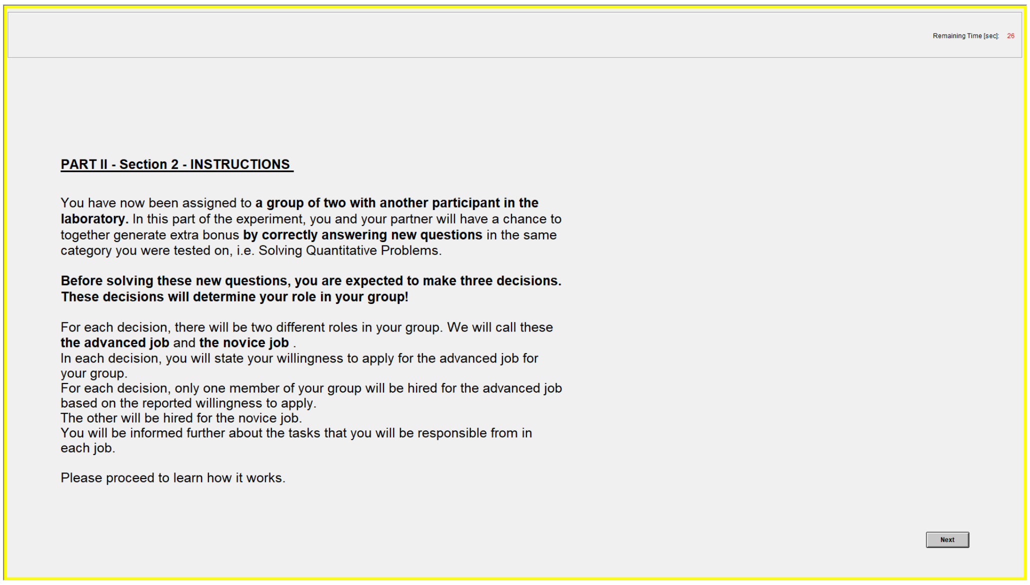


Figure 33: Page 32

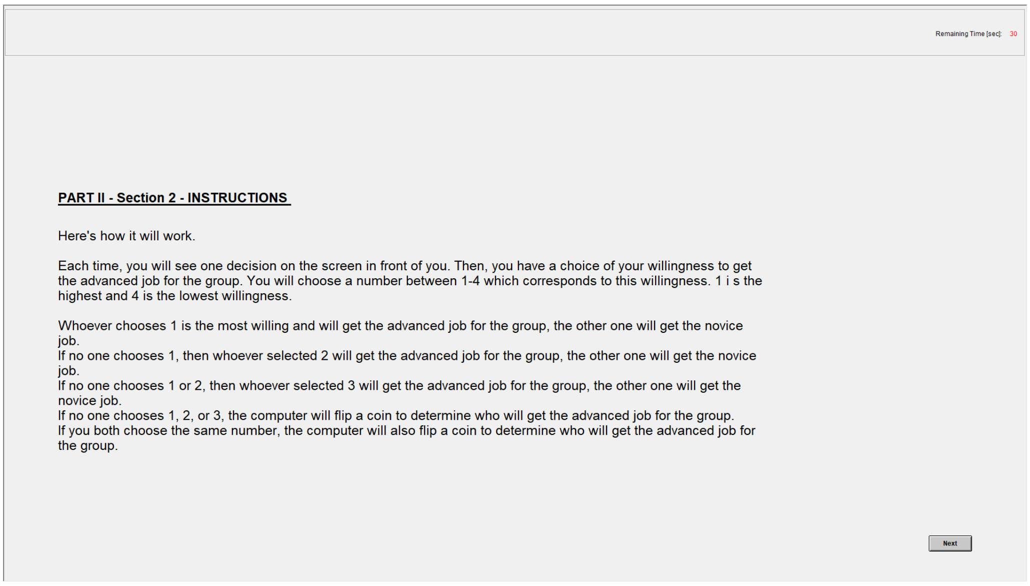


Figure 34: Page 33

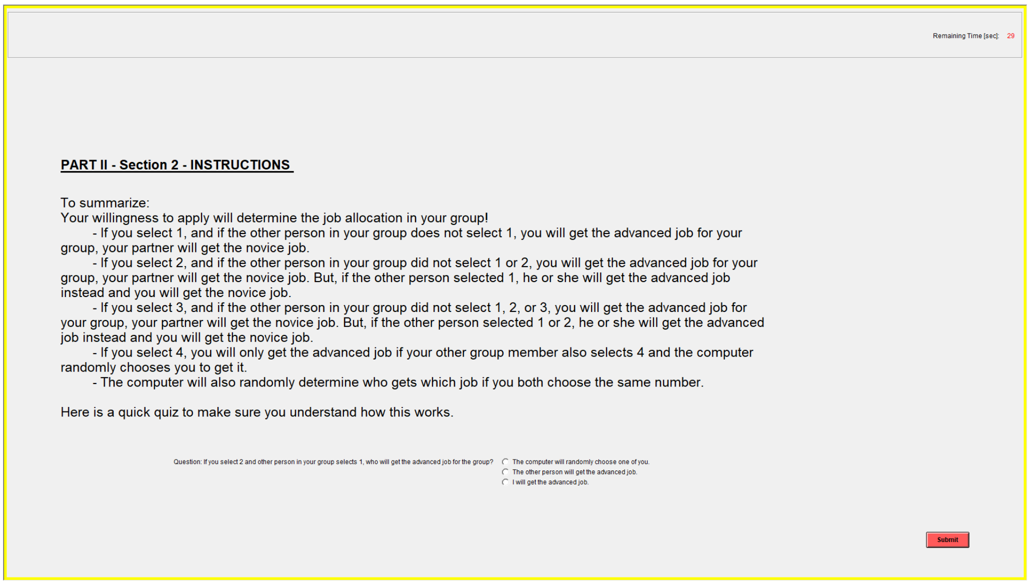


Figure 35: Page 34

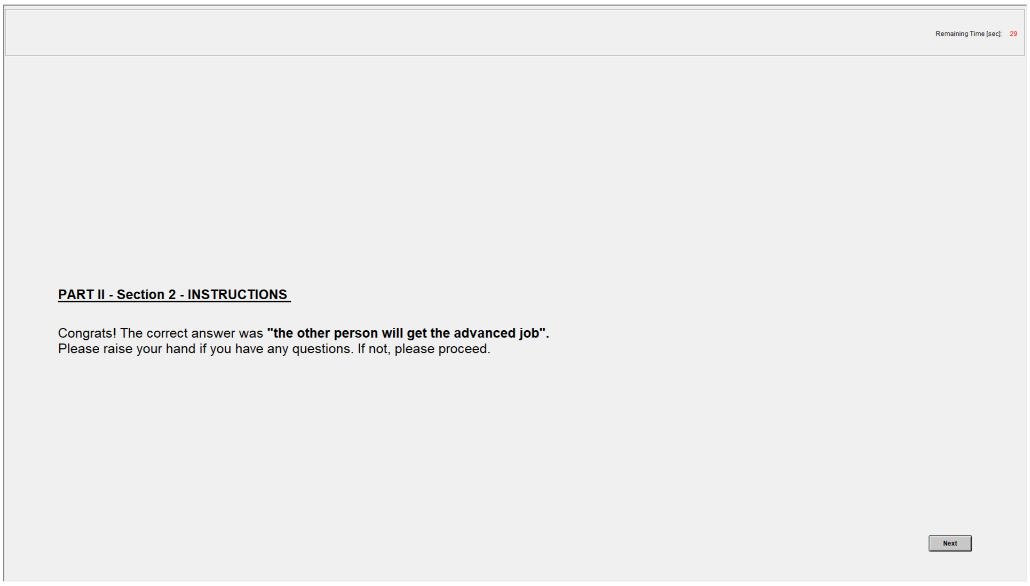


Figure 36: Page 35 (Waitingscreen)

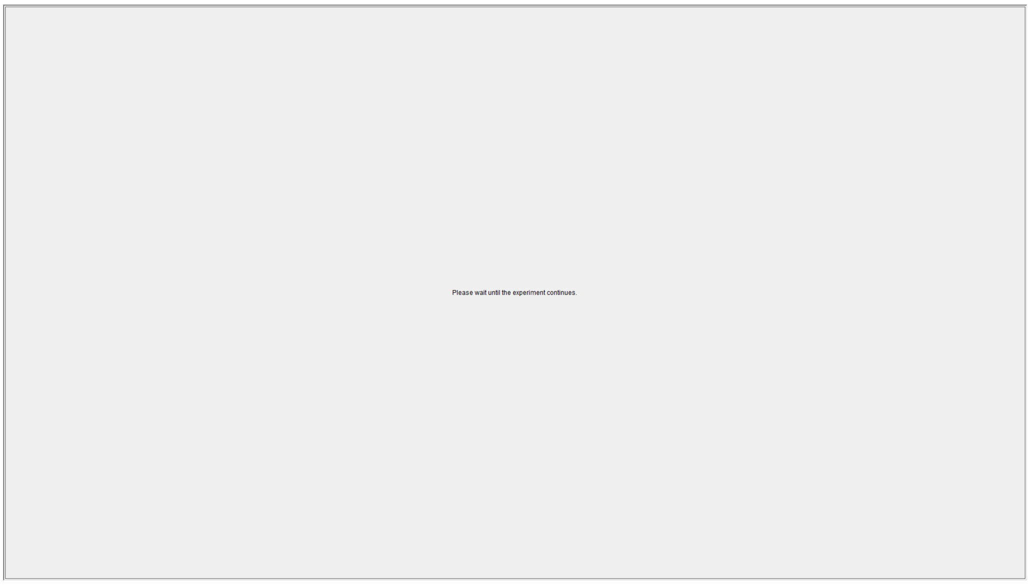


Figure 37: Page 36

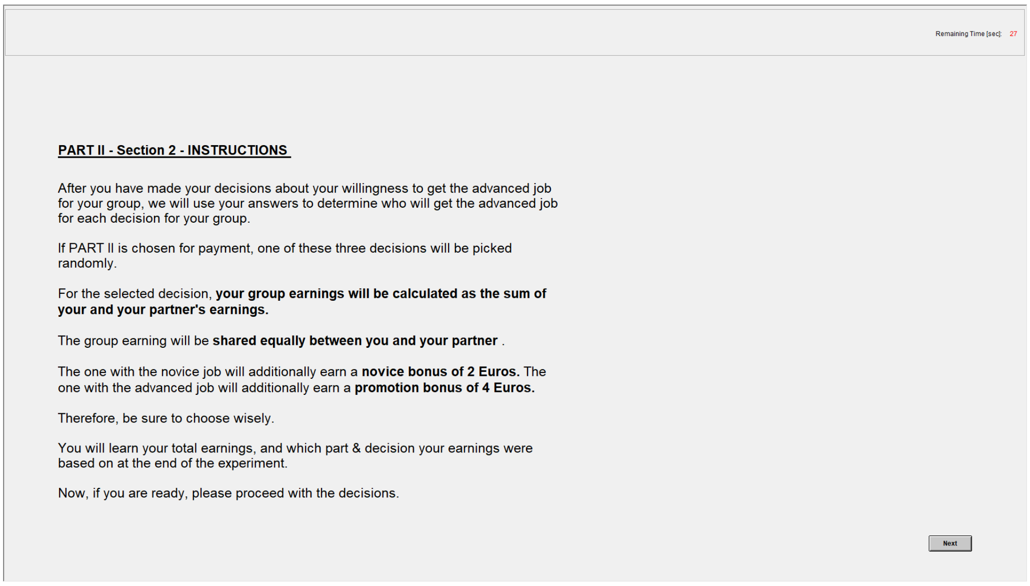


Figure 38: Page 37

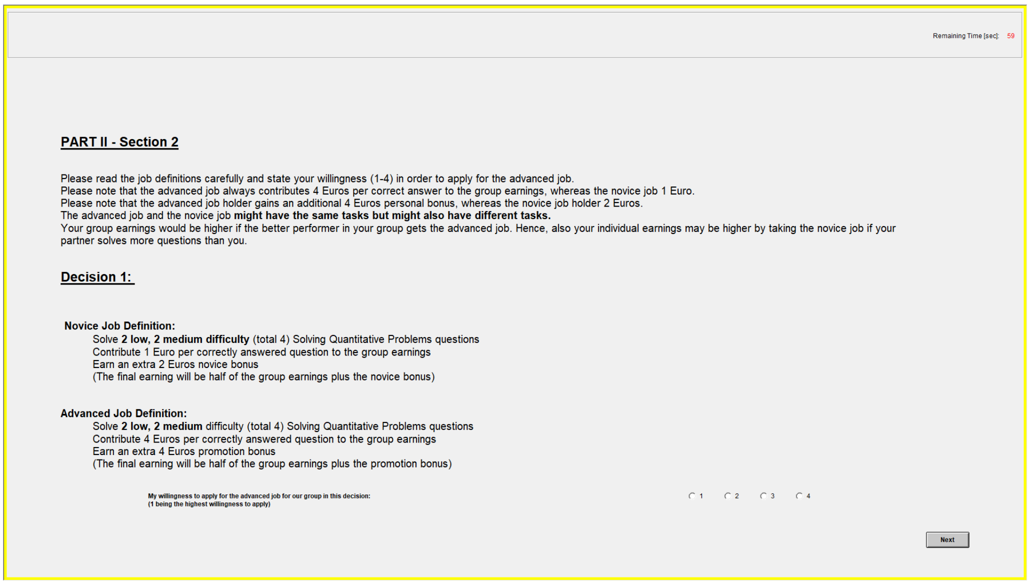


Figure 39: Page 38

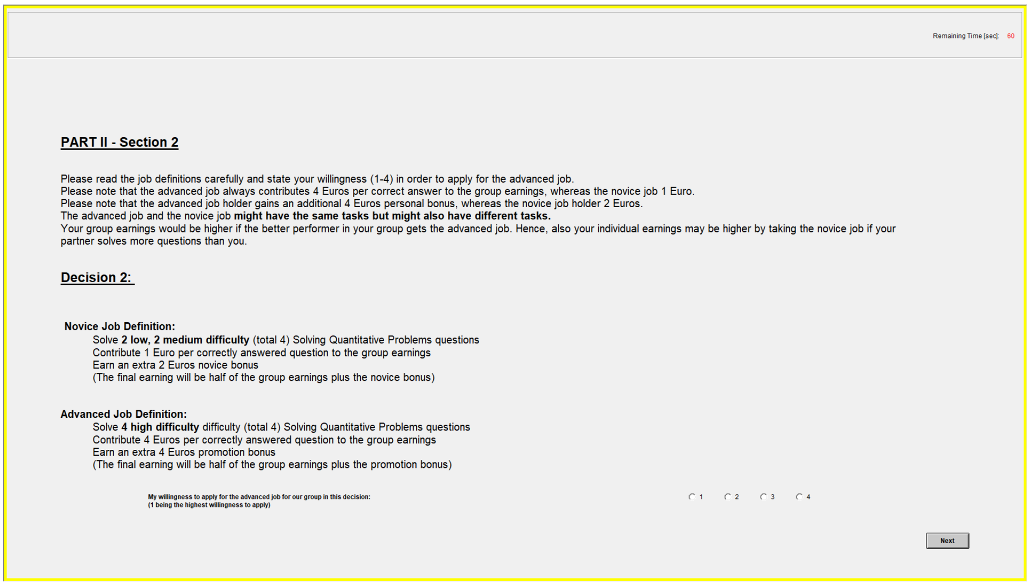


Figure 40: Page 39

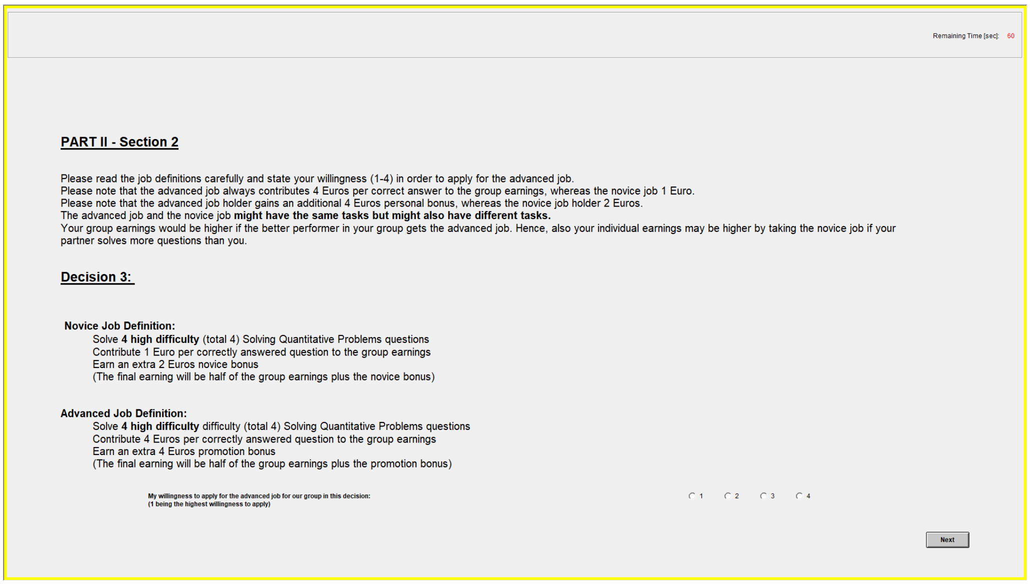


Figure 41: Page 40

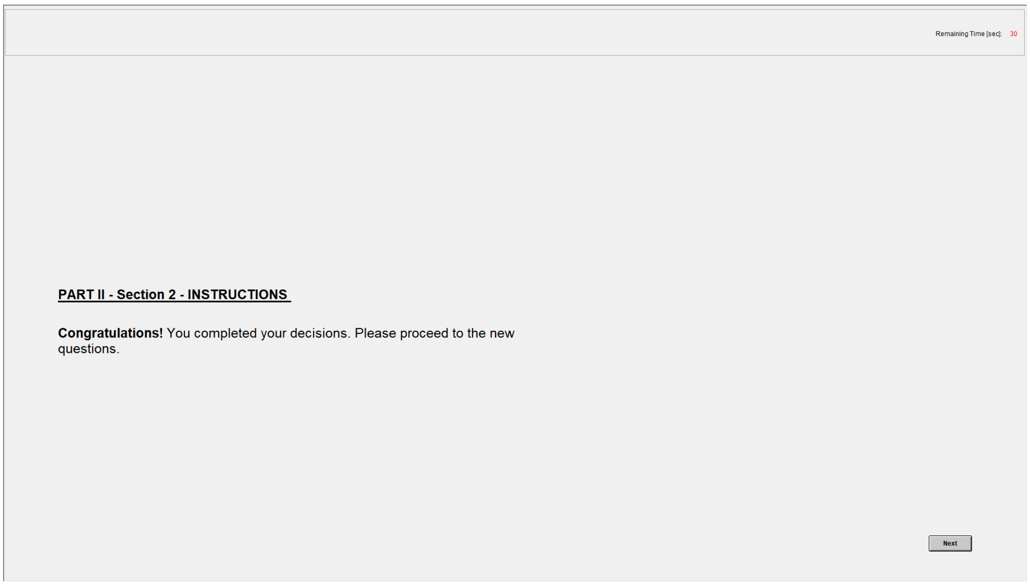


Figure 42: Page 41

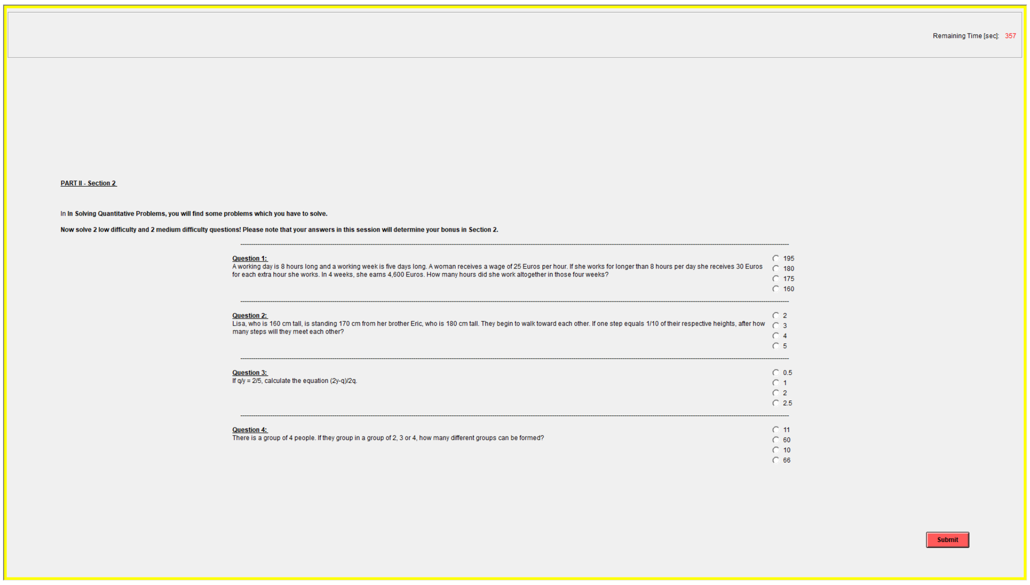


Figure 43: Page 42

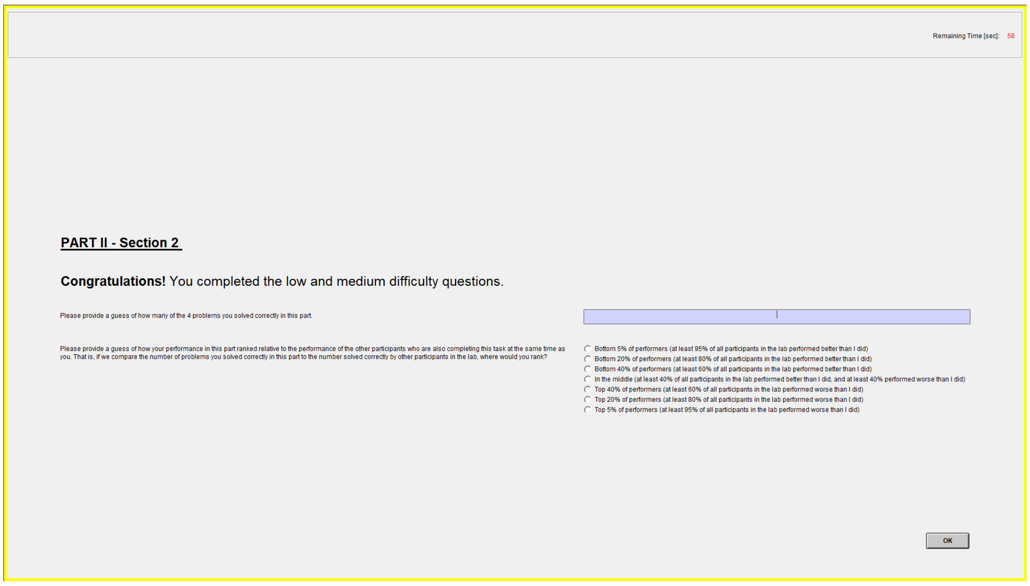


Figure 44: Page 43

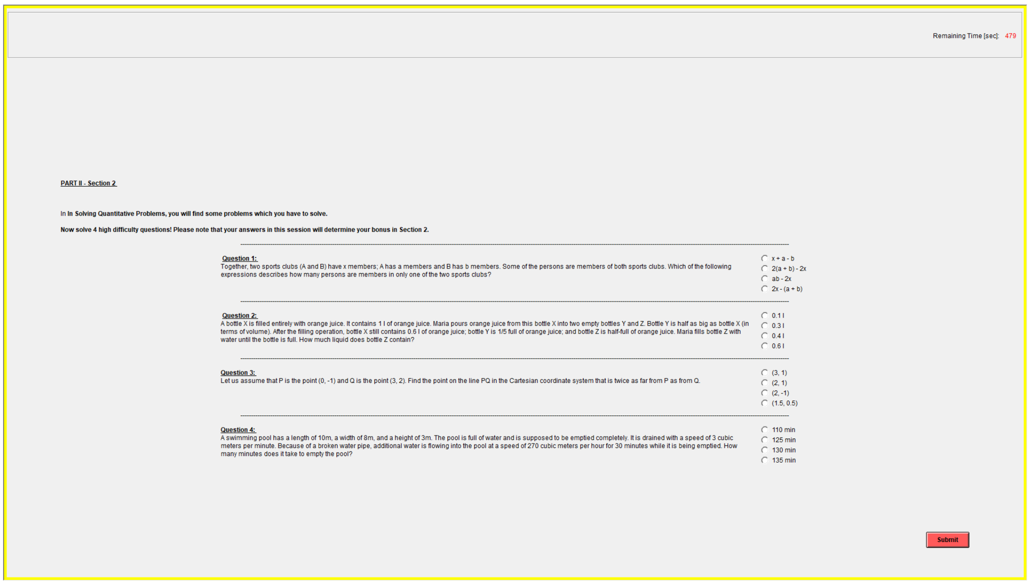


Figure 45: Page 44

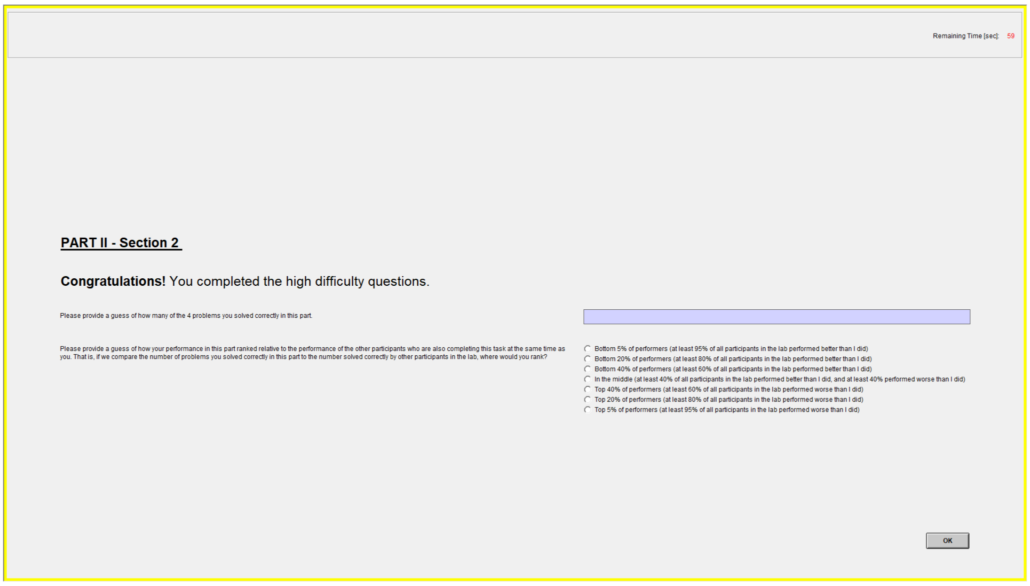


Figure 46: Page 45 (Waitingscreen)

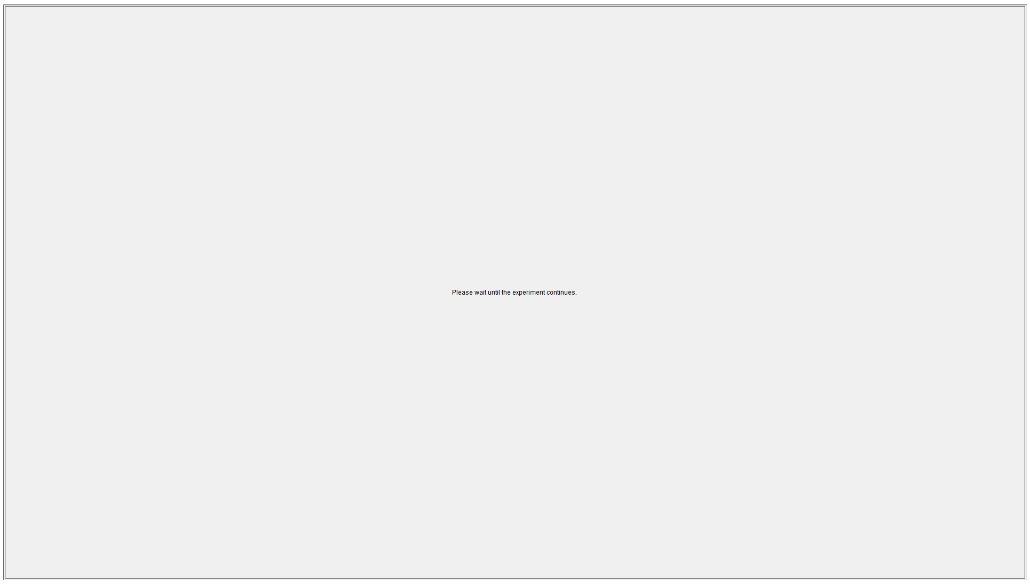


Figure 47: Page 46

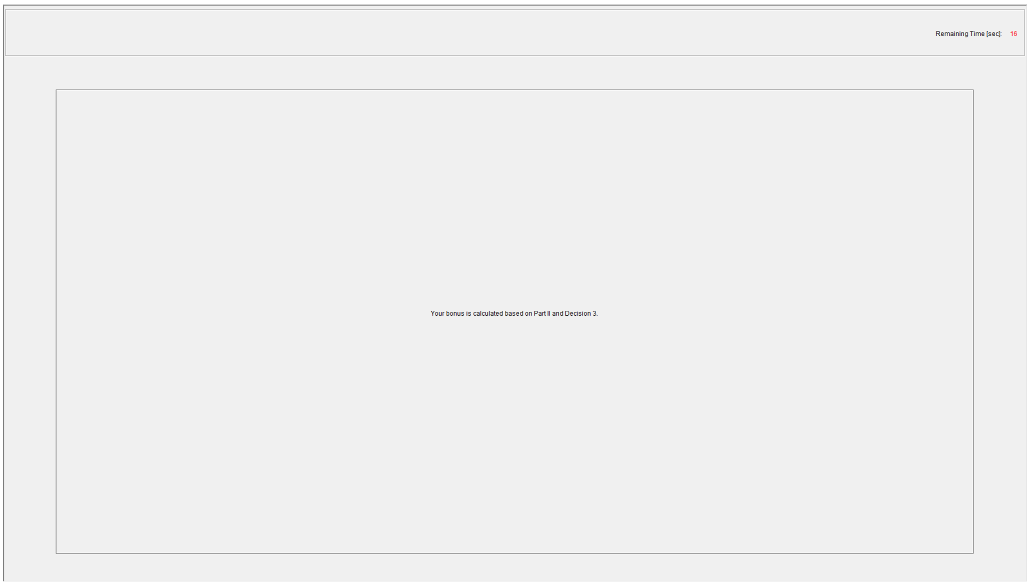


Figure 48: Page 47 (waitingscreen)

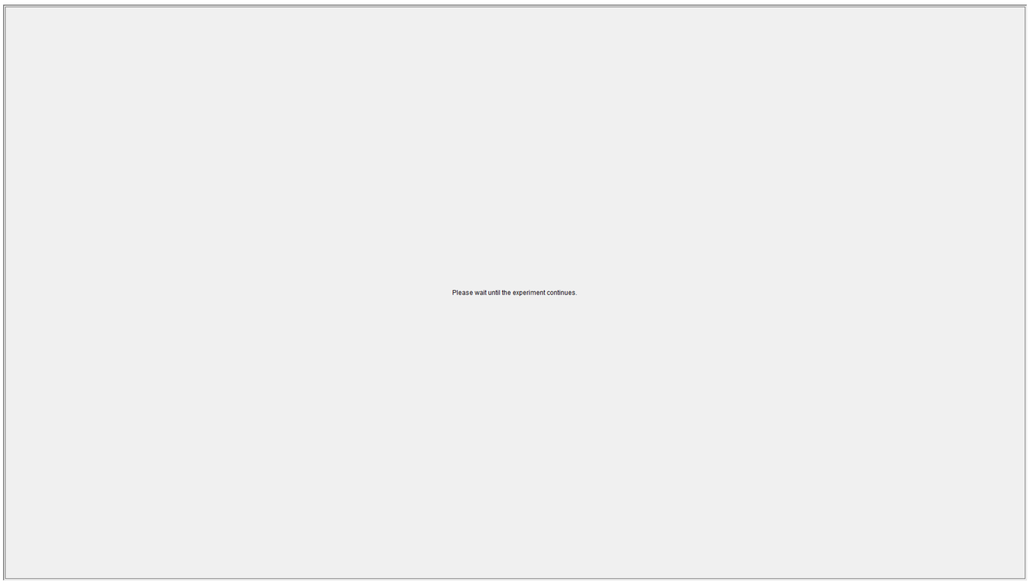


Figure 49: Page 48

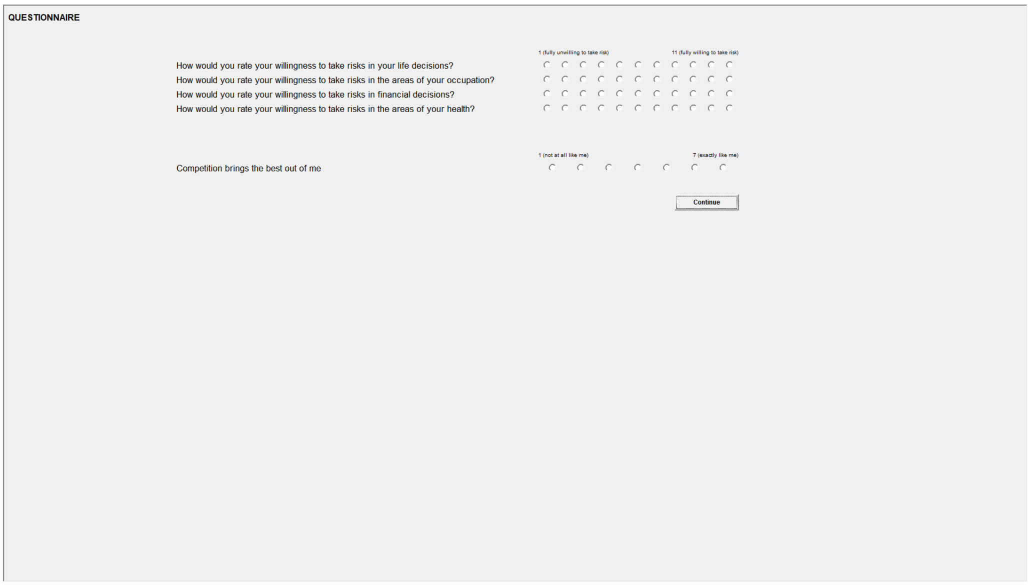
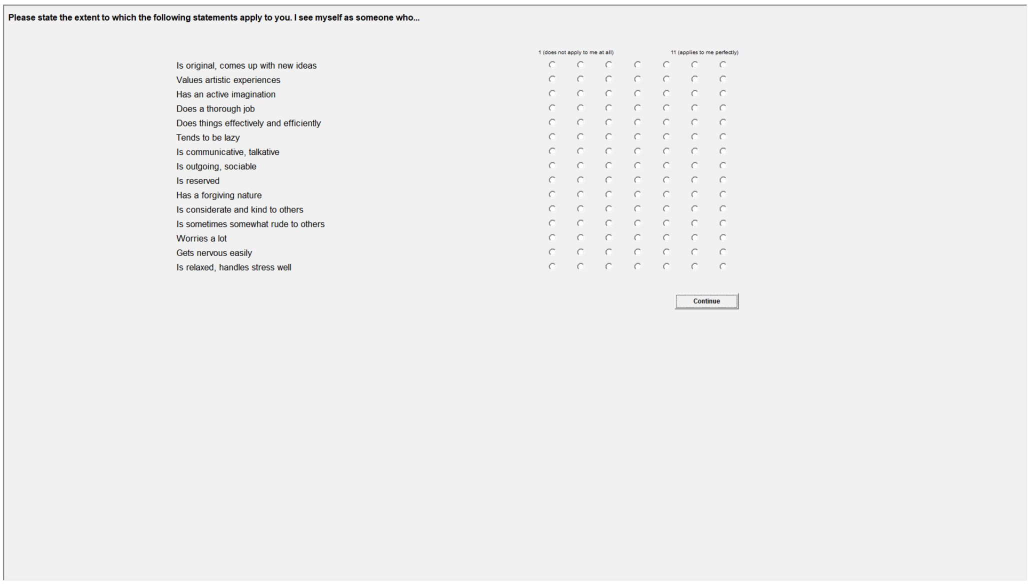


Figure 50: Page 49


Please state the extent to which the following statements apply to you. I see myself as someone who...

1 (does not apply to me at all)
11 (applies to me perfectly)

Is original, comes up with new ideas
Values artistic experiences
Has an active imagination
Does a thorough job
Does things effectively and efficiently
Tends to be lazy
Is communicative, talkative
Is outgoing, sociable
Is reserved
Has a forgiving nature
Is considerate and kind to others
Is sometimes somewhat rude to others
Worries a lot
Gets nervous easily
Is relaxed, handles stress well

Continue

Figure 51: Page 50


Please answer the following questions:

What is your gender?
Male
Female
Other

Where would you put yourself on this scale?
1 (very masculine)
7 (very feminine)

What is your age?

What is your marital status?
Single
Married
In a Domestic Partnership
Divorced
Widowed
Prefer Not to Answer

What is your religion?
Protestant
Catholic
Jewish
Muslim
Hindu
Agnostic
No religion
Other

What is your primary language spoken in your childhood home?
German
English
Turkish
Chinese
Vietnamese
Polish
Other

What is your employment status?
Employed full-time (40 hours a week)
Employed part-time (less than 40 hours a week)
Unemployed (currently looking for work)
Unemployed (not looking for work)
Student
Self-employed
Unable to work

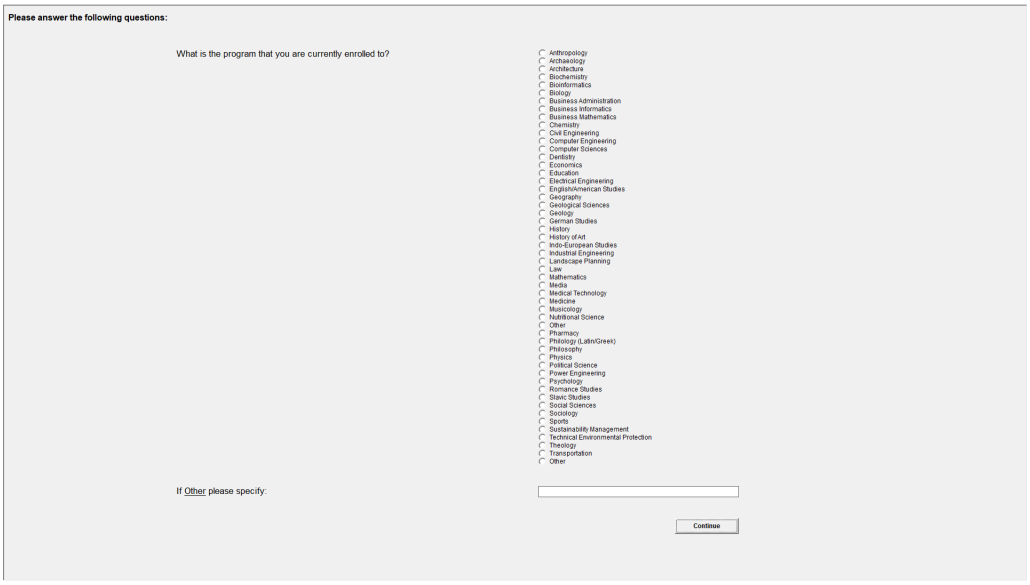
What is the highest level of education that you have completed?
Primary school
High school
Bachelor degree
Master degree
PhD degree

What is the highest level of education that your mother has completed?
Primary school
High school
Bachelor degree
Master degree
PhD degree

What is the highest level of education that your father has completed?
Primary school
High school
Bachelor degree
Master degree
PhD degree

Continue

Figure 52: Page 51



Please answer the following questions:

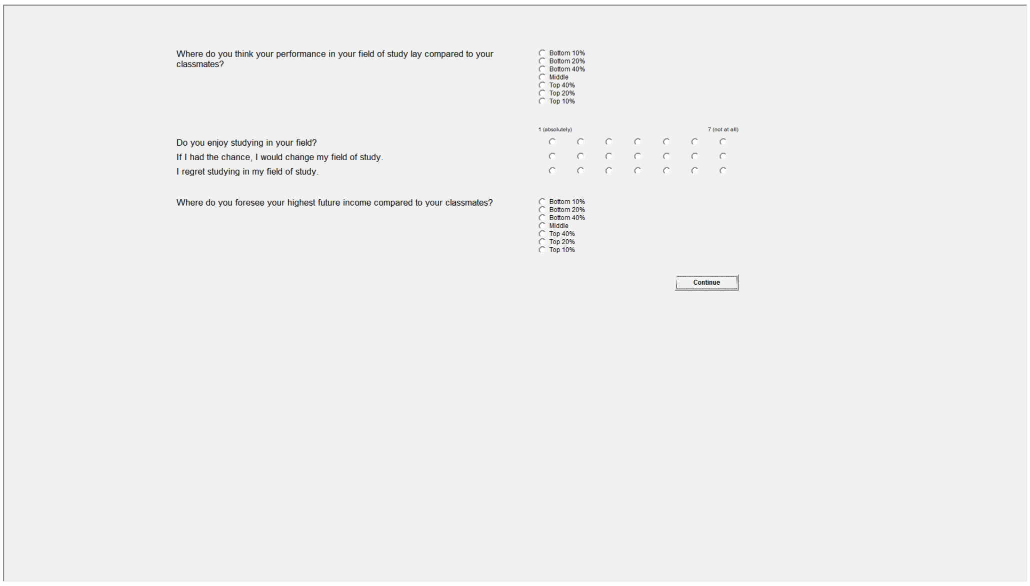
What is the program that you are currently enrolled to?

- Anthropology
- Archaeology
- Architecture
- Biochemistry
- Bioinformatics
- Biology
- Business Administration
- Business Informatics
- Business Mathematics
- Chemistry
- Civil Engineering
- Computer Engineering
- Computer Sciences
- Dentistry
- Economics
- Education
- Electrical Engineering
- English/American Studies
- Geography
- Geological Sciences
- Geology
- German Studies
- History
- History of Art
- Indo-European Studies
- Industrial Engineering
- Landscape Planning
- Law
- Mathematics
- Media
- Medical Technology
- Medicine
- Musicology
- Nutritional Science
- Other
- Pharmacy
- Philology (Latin/Greek)
- Philosophy
- Physics
- Political Science
- Power Engineering
- Psychology
- Romance Studies
- Slavic Studies
- Social Sciences
- Sociology
- Sports
- Sustainability Management
- Technical Environmental Protection
- Theology
- Transportation
- Other

If Other please specify:

Continue

Figure 53: Page 52



Where do you think your performance in your field of study lay compared to your classmates?

- Bottom 10%
- Bottom 20%
- Bottom 40%
- Middle
- Top 40%
- Top 20%
- Top 10%

1 (absolutely) … 7 (not at all)

Do you enjoy studying in your field?

If I had the chance, I would change my field of study.

I regret studying in my field of study.

Where do you foresee your highest future income compared to your classmates?

- Bottom 10%
- Bottom 20%
- Bottom 40%
- Middle
- Top 40%
- Top 20%
- Top 10%

Continue

Figure 54: Page 53

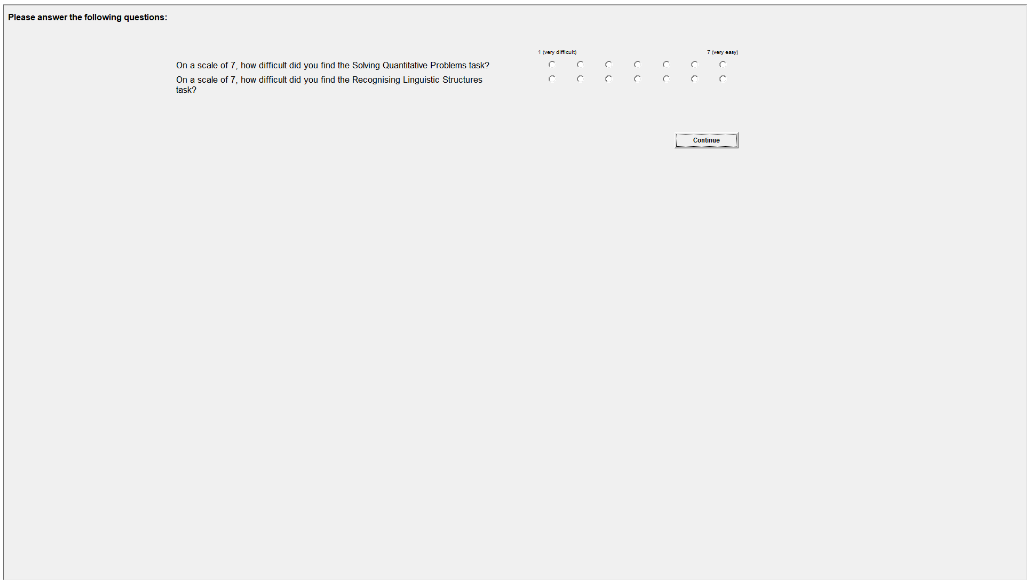


Figure 55: Page 54

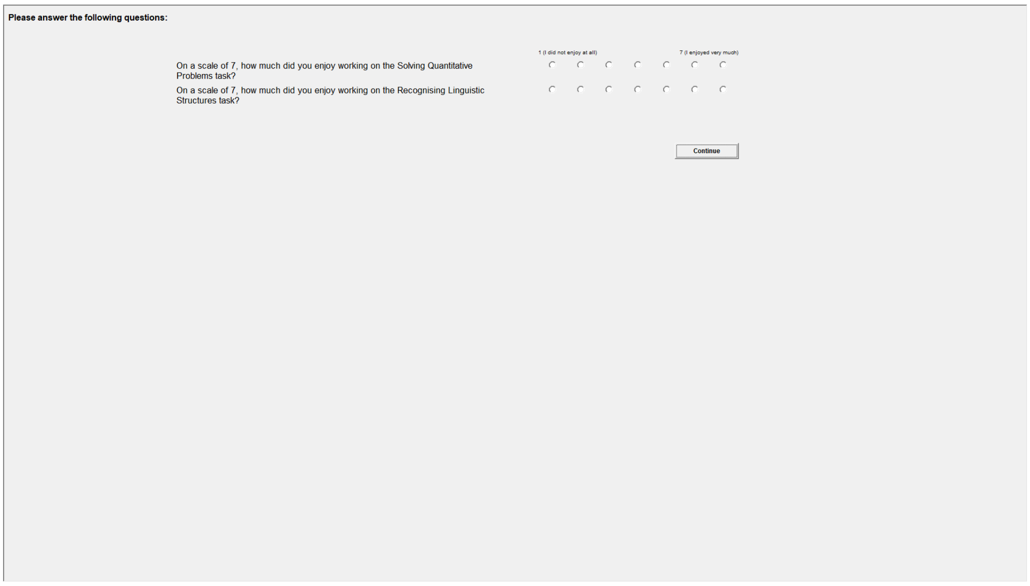


Figure 56: Page 55

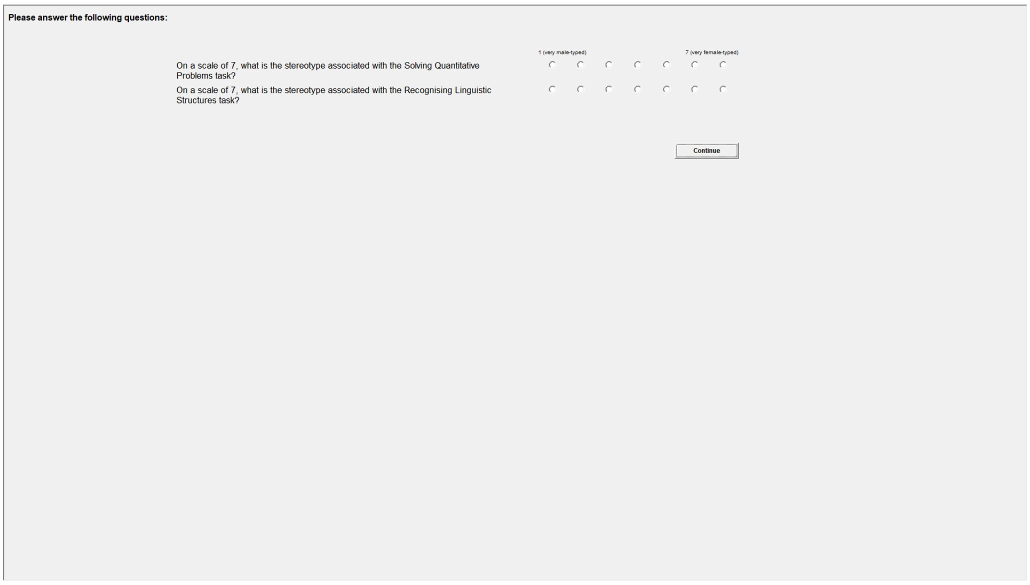


Figure 57: Page 56

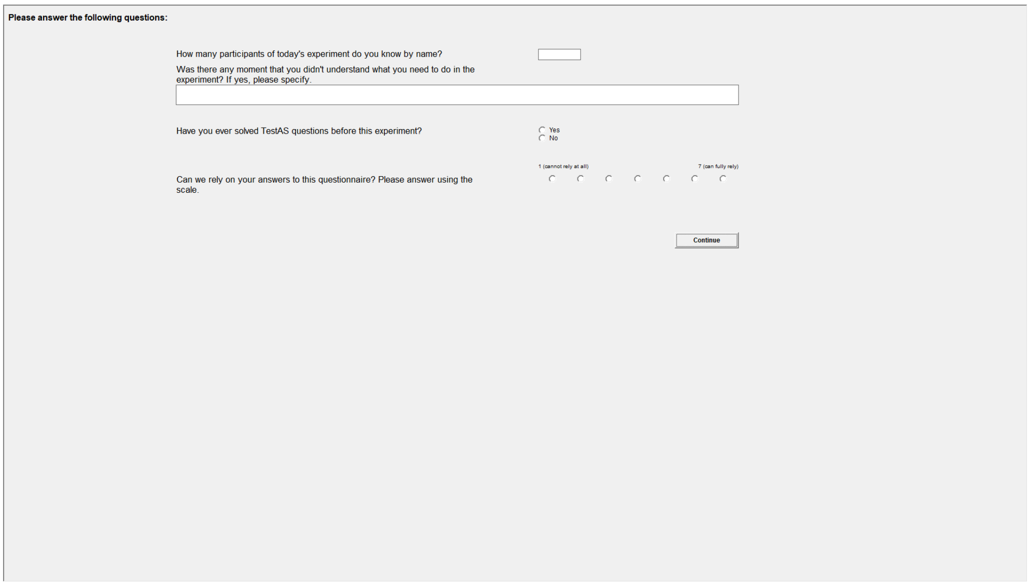
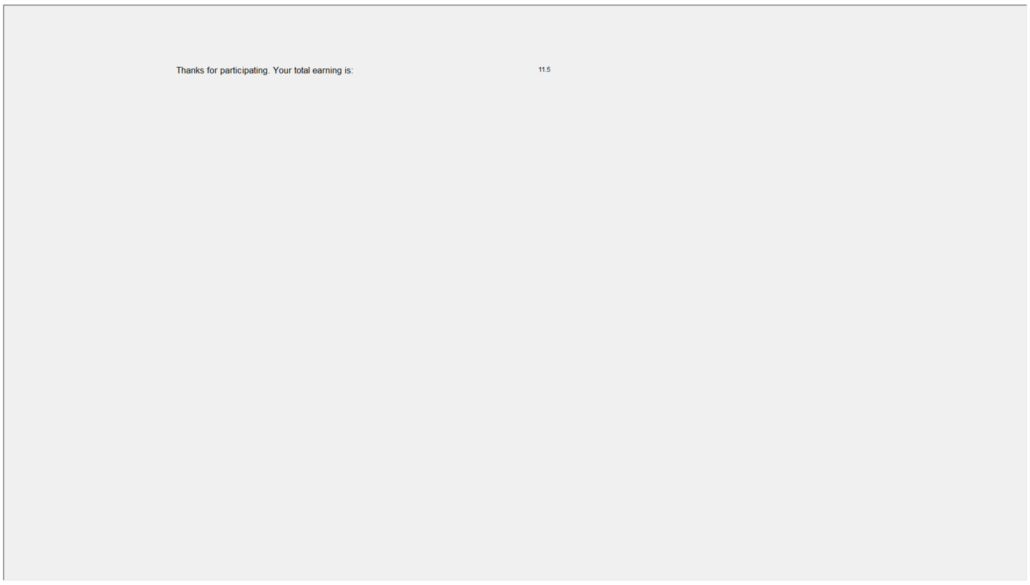
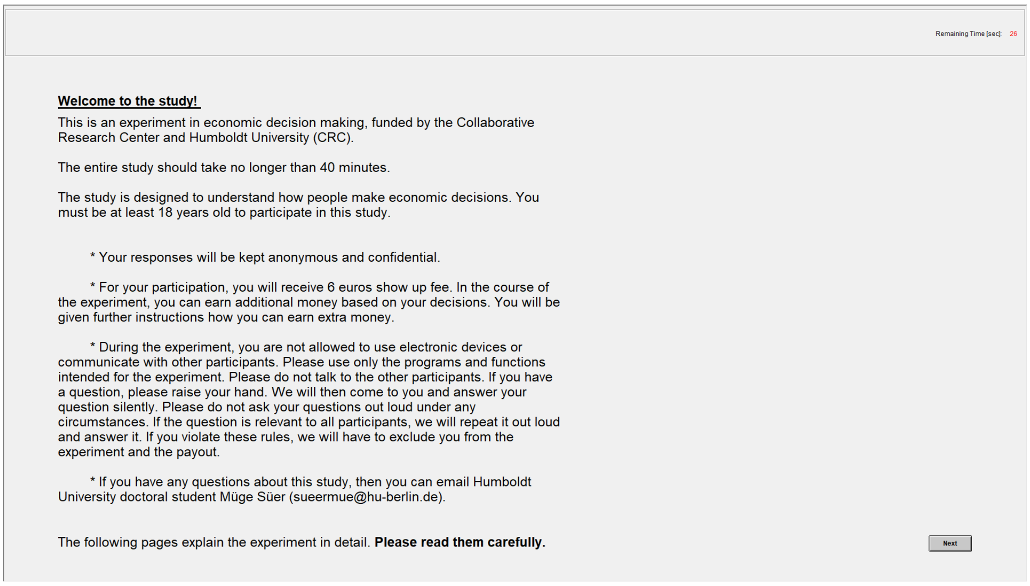


Figure 58: Page 57



A.3.2 Experiment 2

Figure 59: Page 1



Remaining Time [sec]: 26

Welcome to the study!

This is an experiment in economic decision making, funded by the Collaborative Research Center and Humboldt University (CRC).

The entire study should take no longer than 40 minutes.

The study is designed to understand how people make economic decisions. You must be at least 18 years old to participate in this study.

* Your responses will be kept anonymous and confidential.

* For your participation, you will receive 6 euros show up fee. In the course of the experiment, you can earn additional money based on your decisions. You will be given further instructions how you can earn extra money.

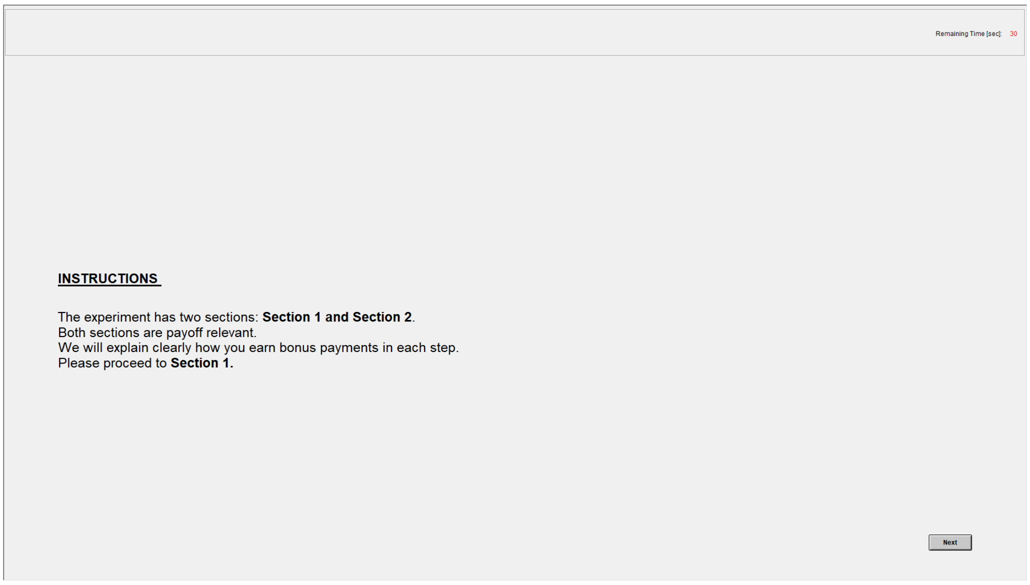
* During the experiment, you are not allowed to use electronic devices or communicate with other participants. Please use only the programs and functions intended for the experiment. Please do not talk to the other participants. If you have a question, please raise your hand. We will then come to you and answer your question silently. Please do not ask your questions out loud under any circumstances. If the question is relevant to all participants, we will repeat it out loud and answer it. If you violate these rules, we will have to exclude you from the experiment and the payout.

* If you have any questions about this study, then you can email Humboldt University doctoral student Müge Süer (sueermue@hu-berlin.de).

The following pages explain the experiment in detail. **Please read them carefully.**

Next

Figure 60: Page 2



Remaining Time [sec]: 30

INSTRUCTIONS

The experiment has two sections: **Section 1 and Section 2**.
Both sections are payoff relevant.
We will explain clearly how you earn bonus payments in each step.
Please proceed to **Section 1.**

Next

Figure 61: Page 3

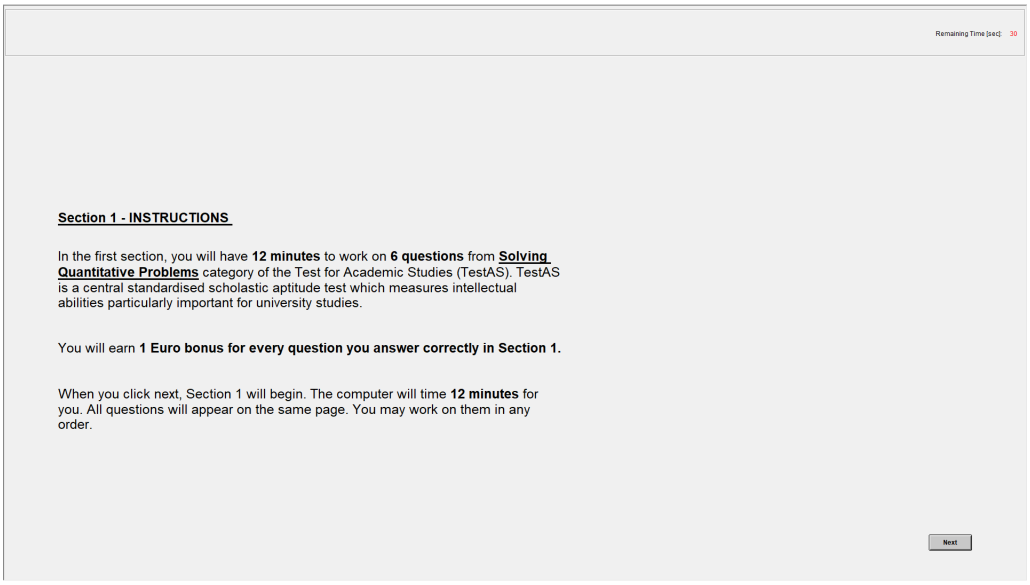


Figure 62: Page 4

Figure 63: Page 5

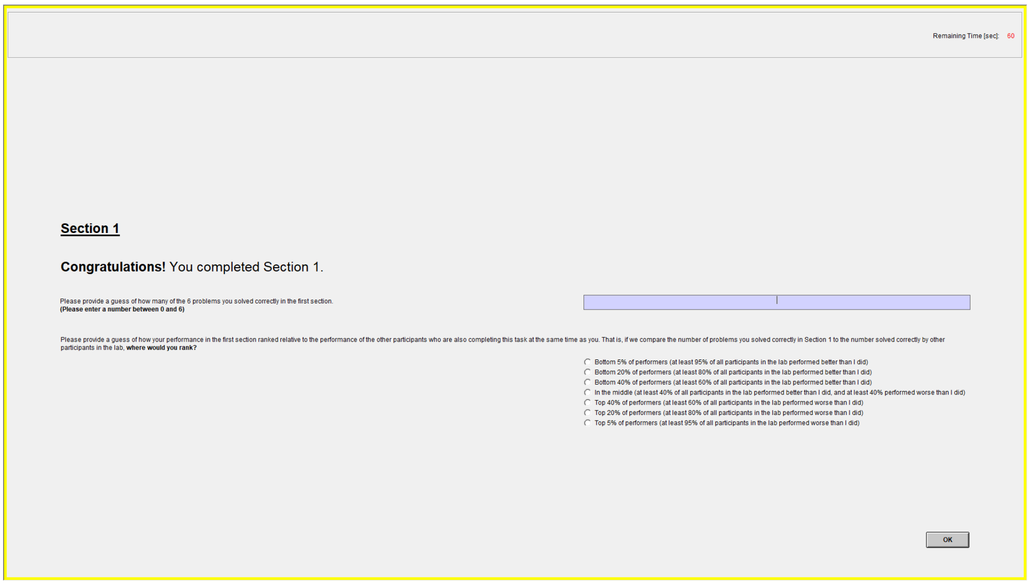


Figure 64: Page 6

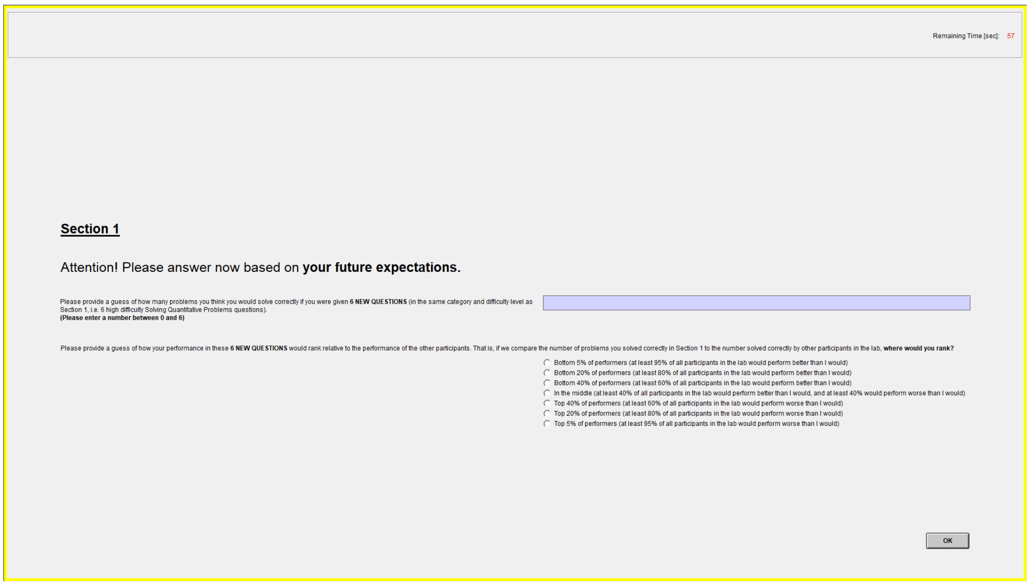


Figure 65: Page 7

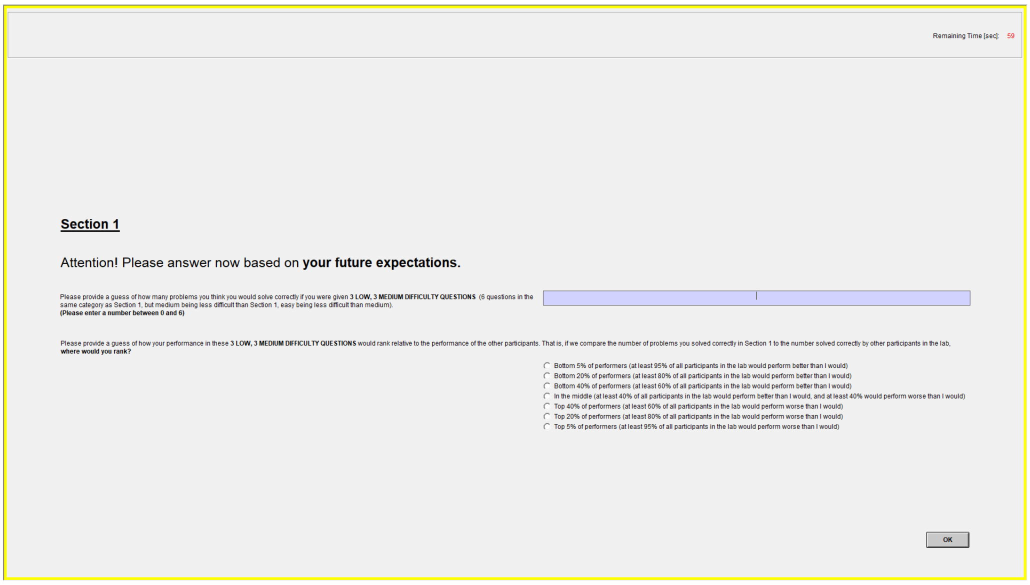


Figure 66: Page 8

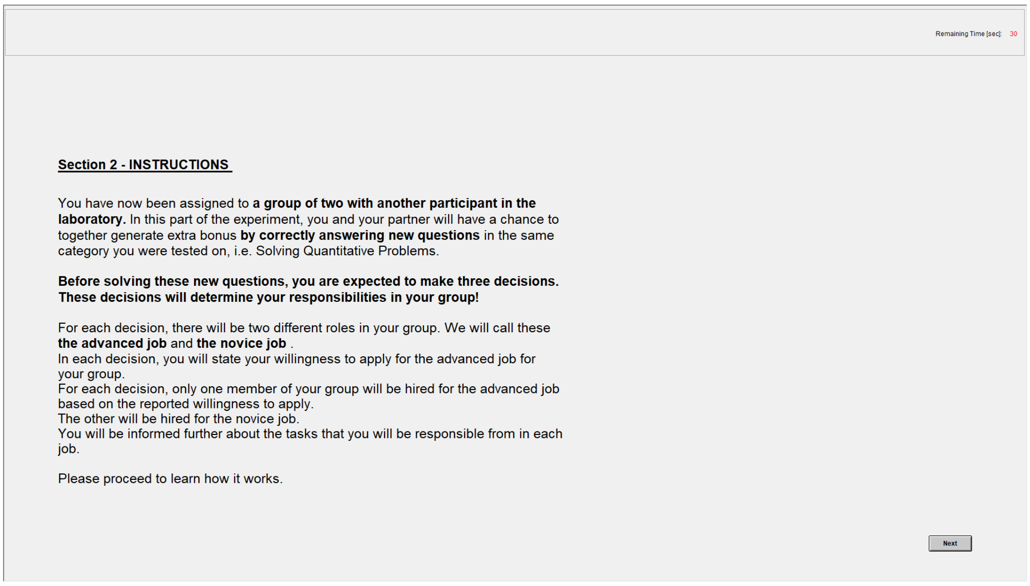


Figure 67: Page 9

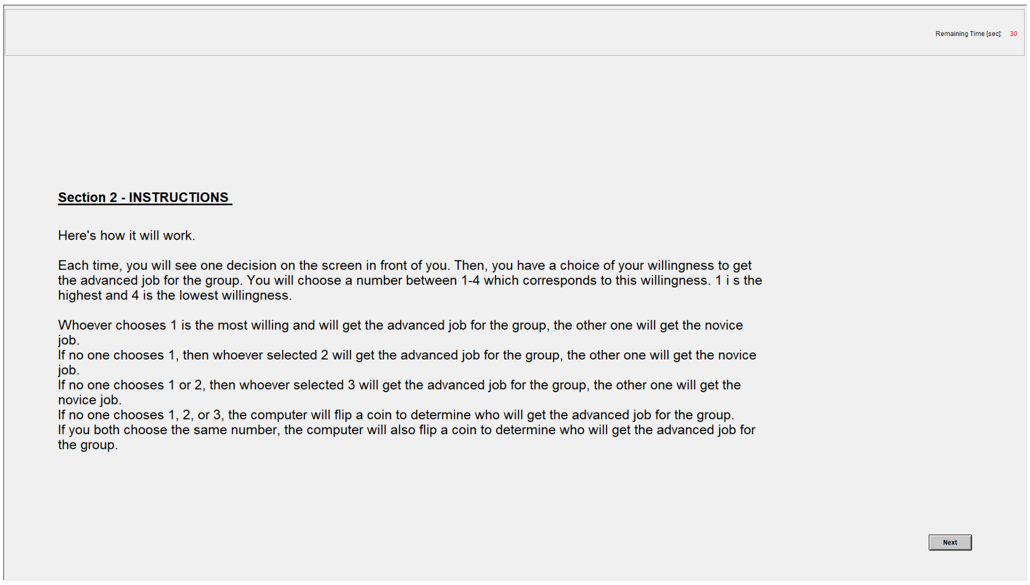


Figure 68: Page 10

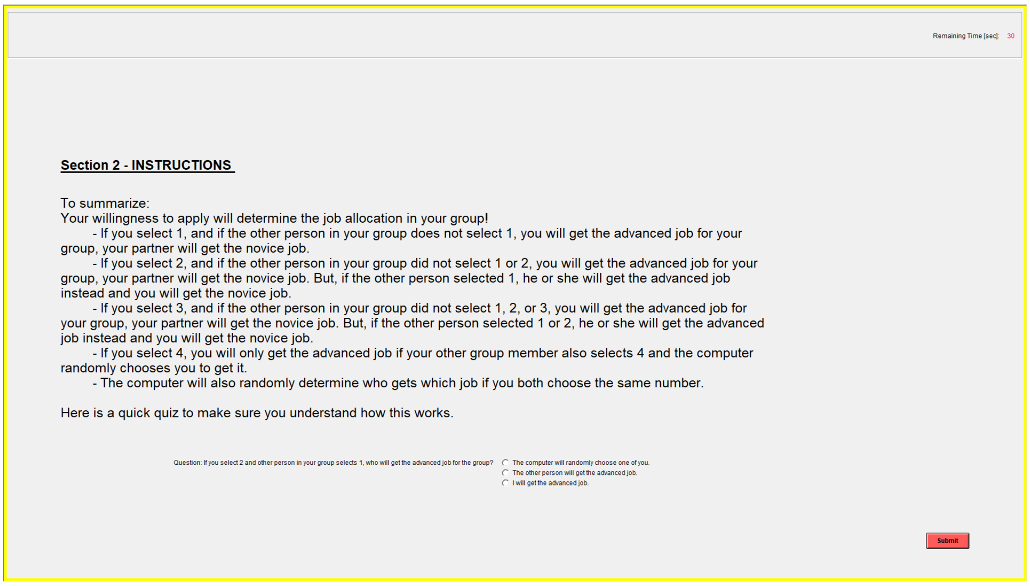


Figure 69: Page 11

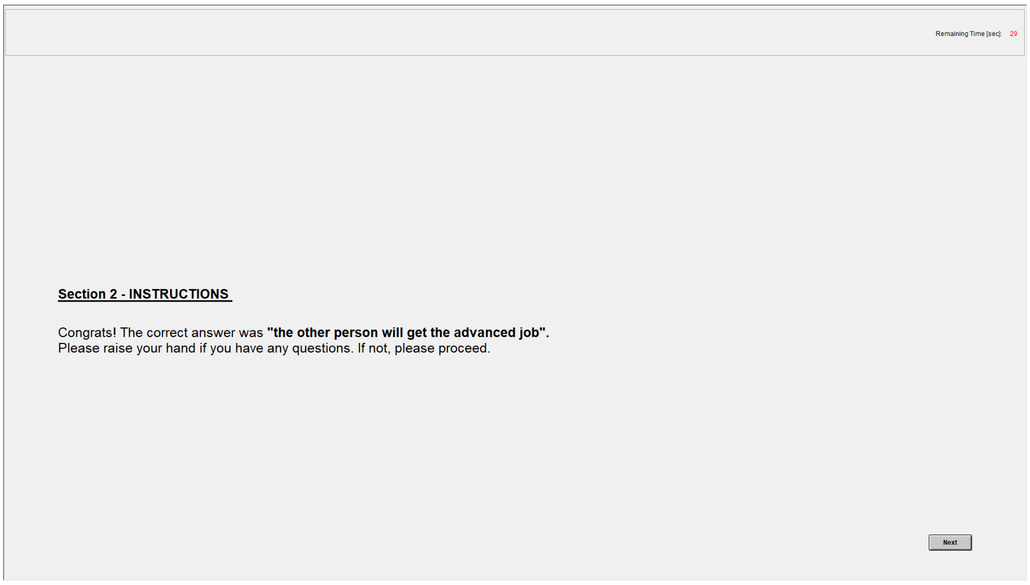


Figure 70: Page 12

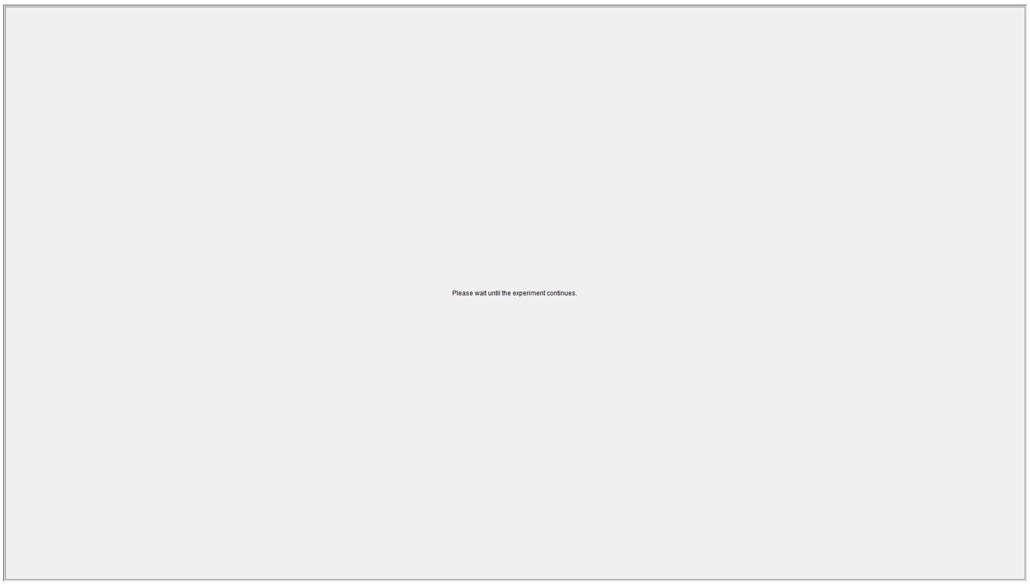
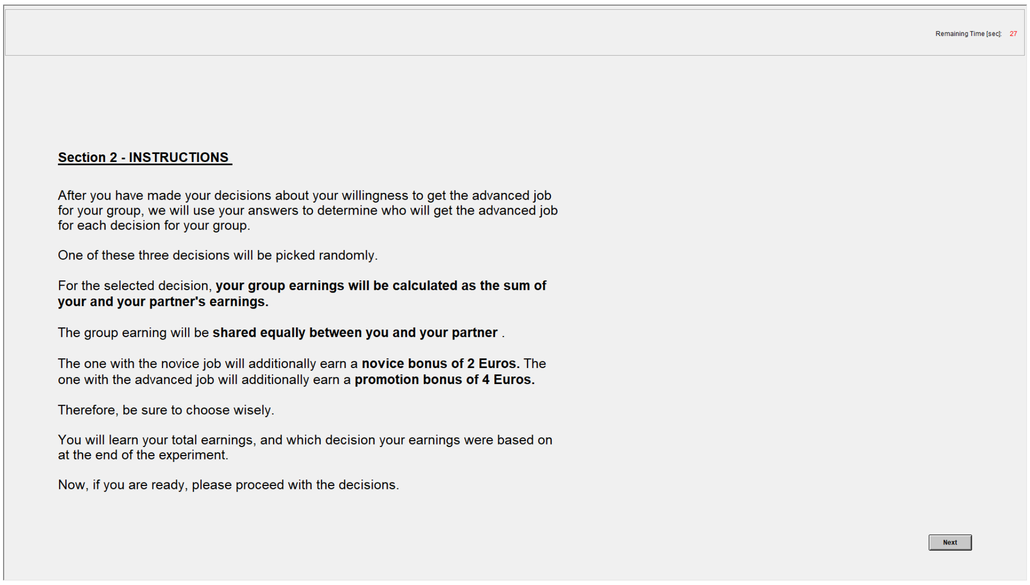


Figure 71: Page 13



Remaining Time [sec]: 27

Section 2 - INSTRUCTIONS

After you have made your decisions about your willingness to get the advanced job for your group, we will use your answers to determine who will get the advanced job for each decision for your group.

One of these three decisions will be picked randomly.

For the selected decision, **your group earnings will be calculated as the sum of your and your partner's earnings.**

The group earning will be **shared equally between you and your partner** .

The one with the novice job will additionally earn a **novice bonus of 2 Euros.** The one with the advanced job will additionally earn a **promotion bonus of 4 Euros.**

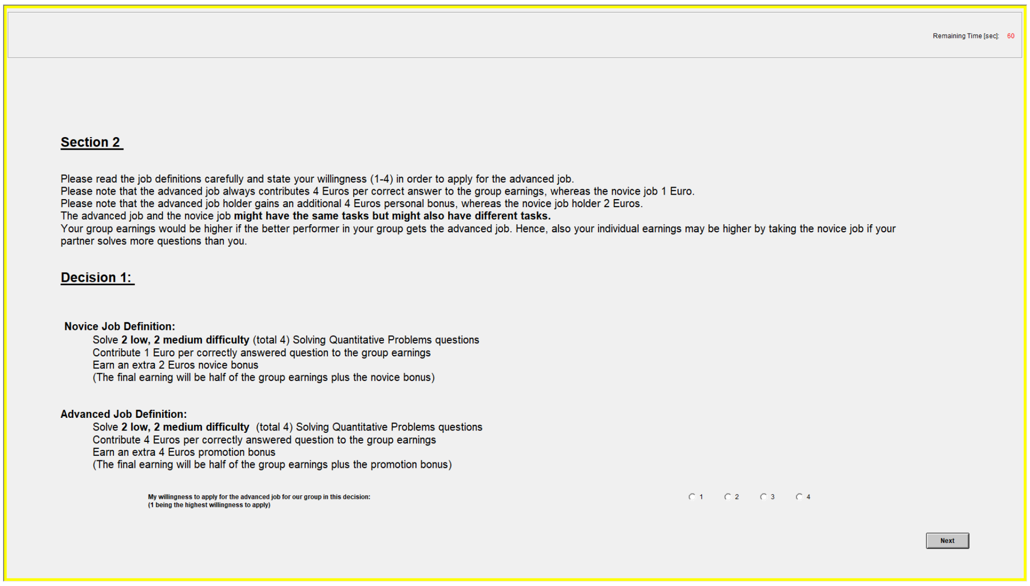
Therefore, be sure to choose wisely.

You will learn your total earnings, and which decision your earnings were based on at the end of the experiment.

Now, if you are ready, please proceed with the decisions.

Next

Figure 72: Page 14



Remaining Time [sec]: 60

Section 2

Please read the job definitions carefully and state your willingness (1-4) in order to apply for the advanced job.
Please note that the advanced job always contributes 4 Euros per correct answer to the group earnings, whereas the novice job 1 Euro.
Please note that the advanced job holder gains an additional 4 Euros personal bonus, whereas the novice job holder 2 Euros.
The advanced job and the novice job **might have the same tasks but might also have different tasks.**
Your group earnings would be higher if the better performer in your group gets the advanced job. Hence, also your individual earnings may be higher by taking the novice job if your partner solves more questions than you.

Decision 1:

Novice Job Definition:
Solve **2 low, 2 medium difficulty** (total 4) Solving Quantitative Problems questions
Contribute 1 Euro per correctly answered question to the group earnings
Earn an extra 2 Euros novice bonus
(The final earning will be half of the group earnings plus the novice bonus)

Advanced Job Definition:
Solve **2 low, 2 medium difficulty** (total 4) Solving Quantitative Problems questions
Contribute 4 Euros per correctly answered question to the group earnings
Earn an extra 4 Euros promotion bonus
(The final earning will be half of the group earnings plus the promotion bonus)

My willingness to apply for the advanced job for our group in this decision:
(1 being the highest willingness to apply) ○ 1 ○ 2 ○ 3 ○ 4

Next

Figure 73: Page 15

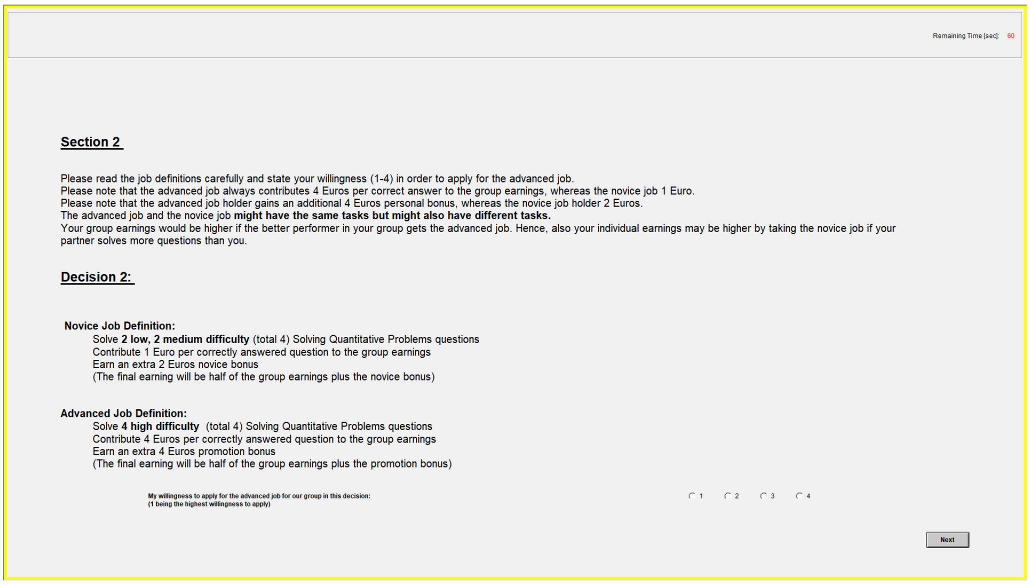


Figure 74: Page 16

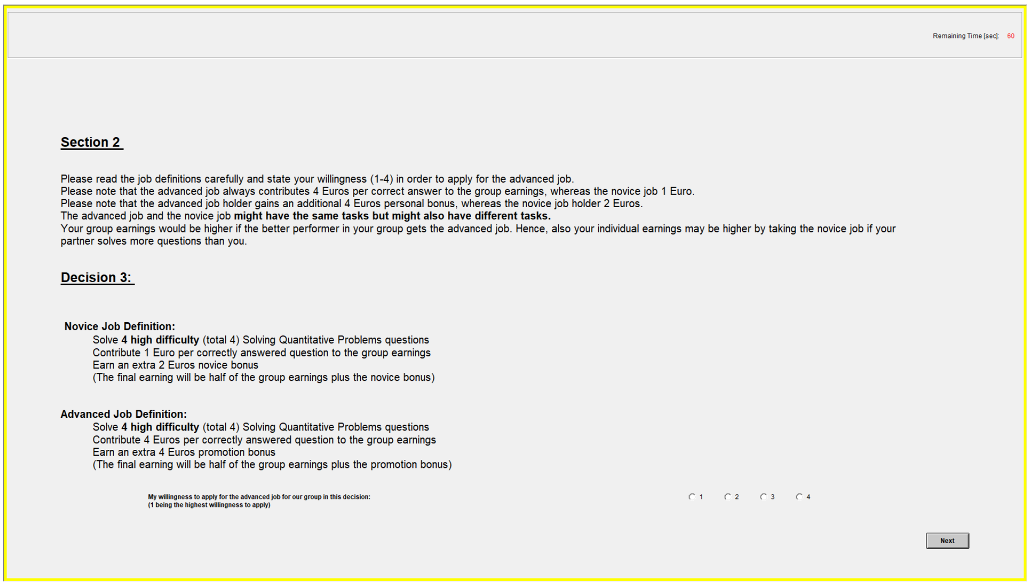


Figure 75: Page 17

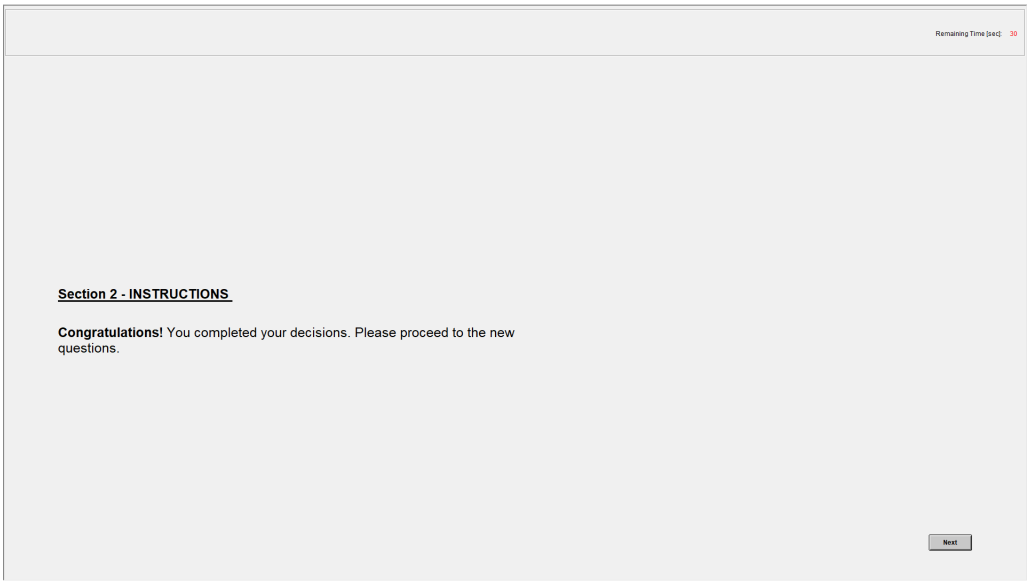


Figure 76: Page 18

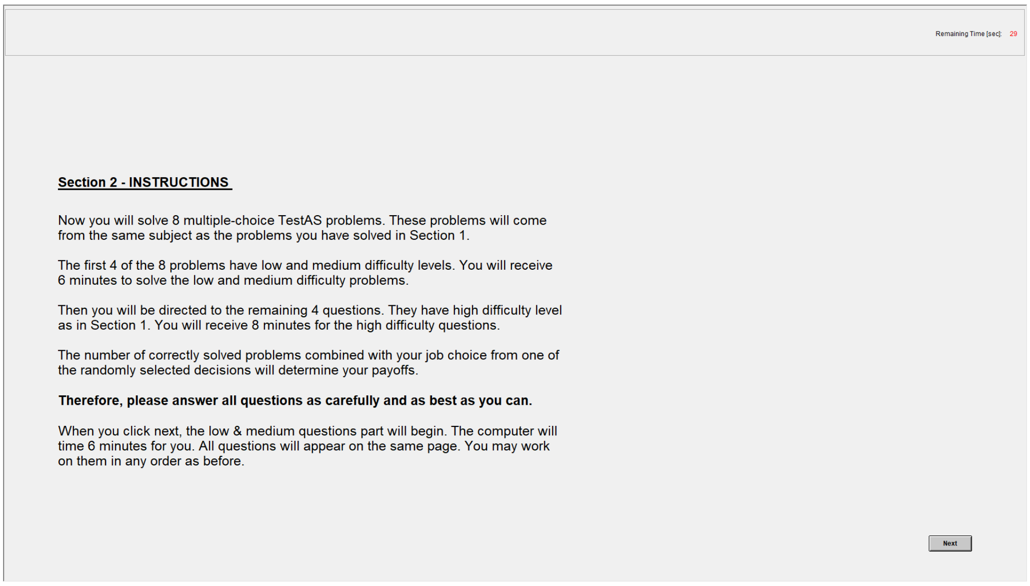


Figure 77: Page 19

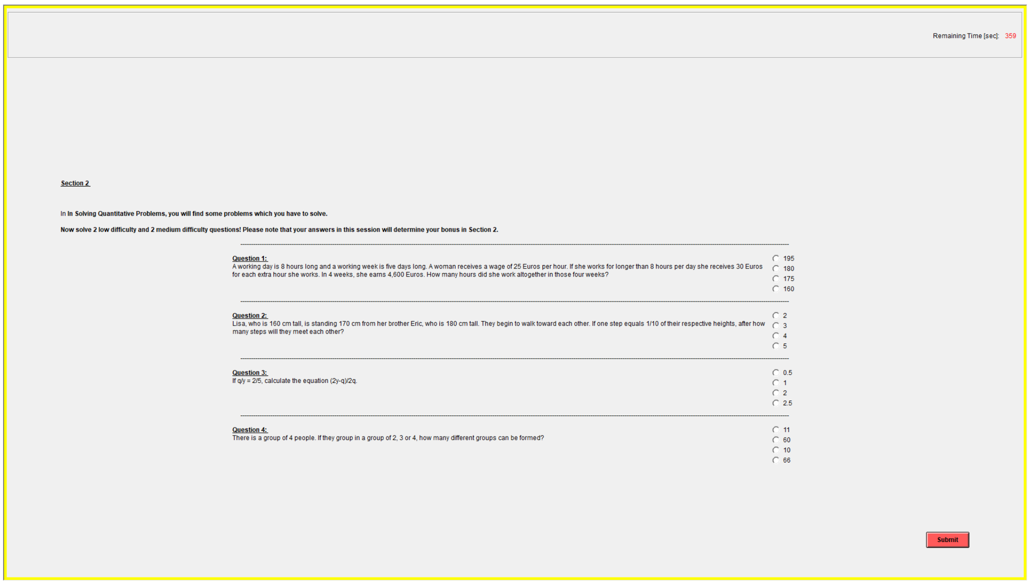


Figure 78: Page 20

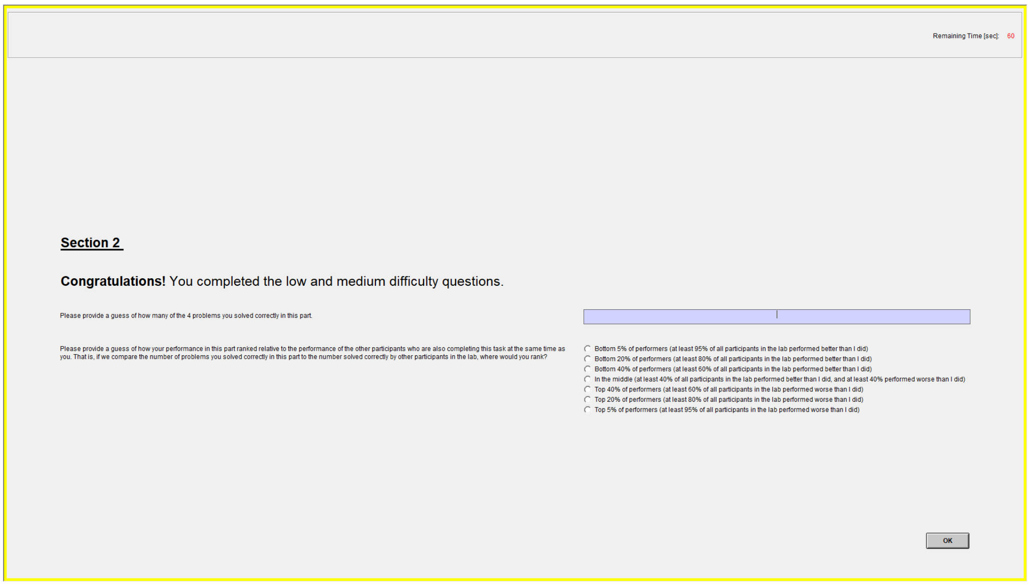


Figure 79: Page 21

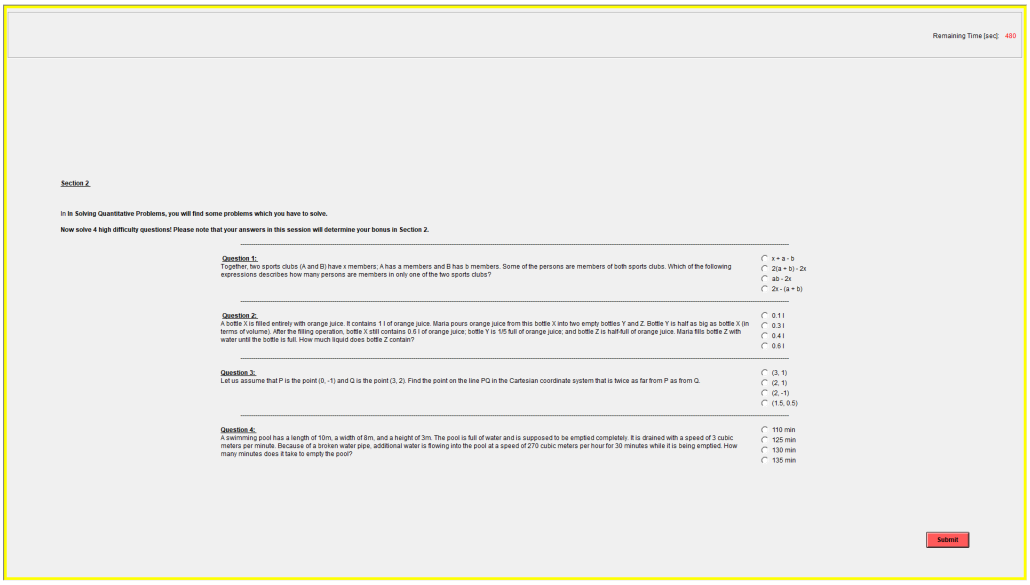


Figure 80: Page 22

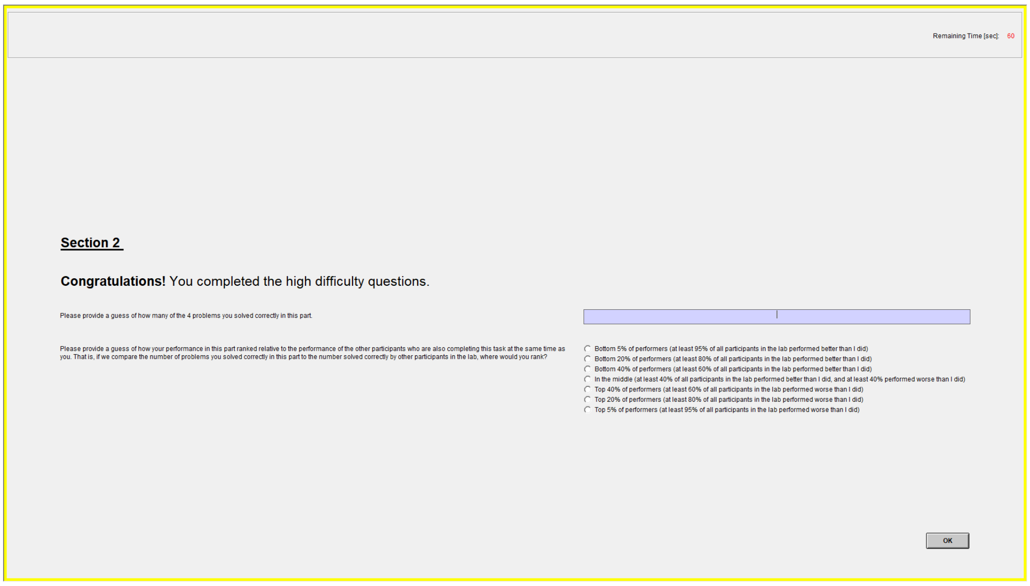


Figure 81: Page 23

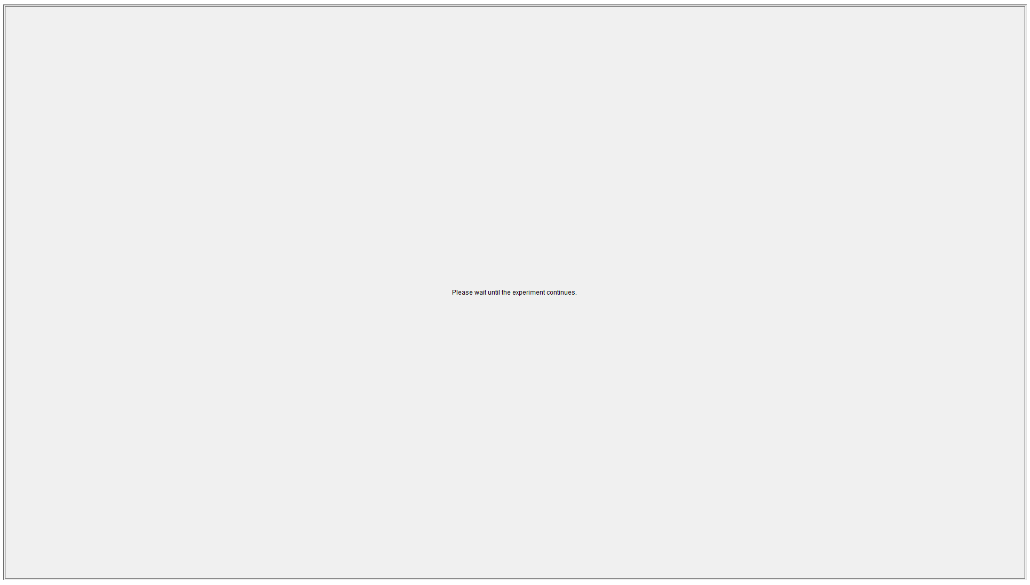


Figure 82: Page 24

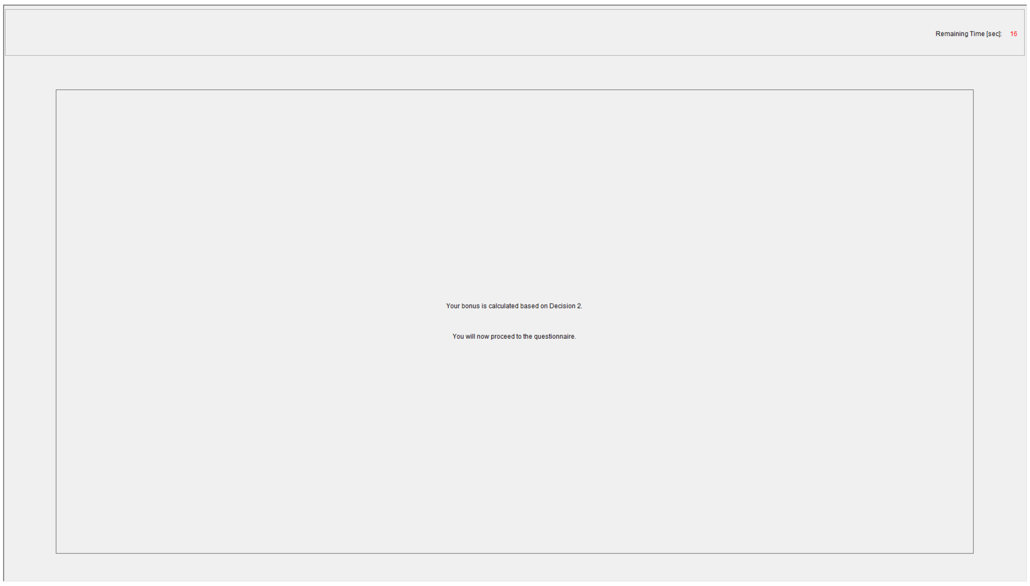


Figure 83: Page 25

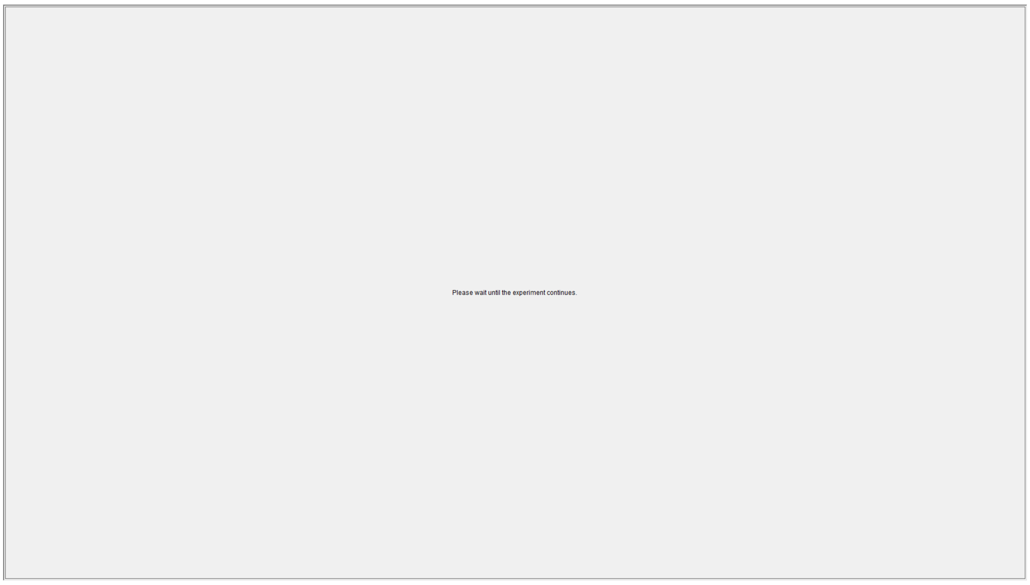


Figure 84: Page 26

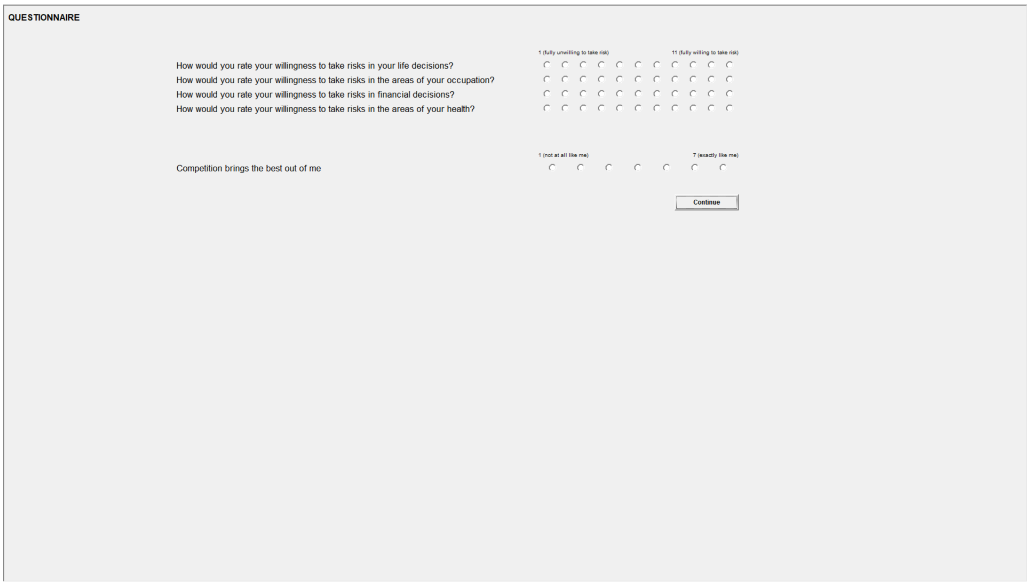


Figure 85: Page 27

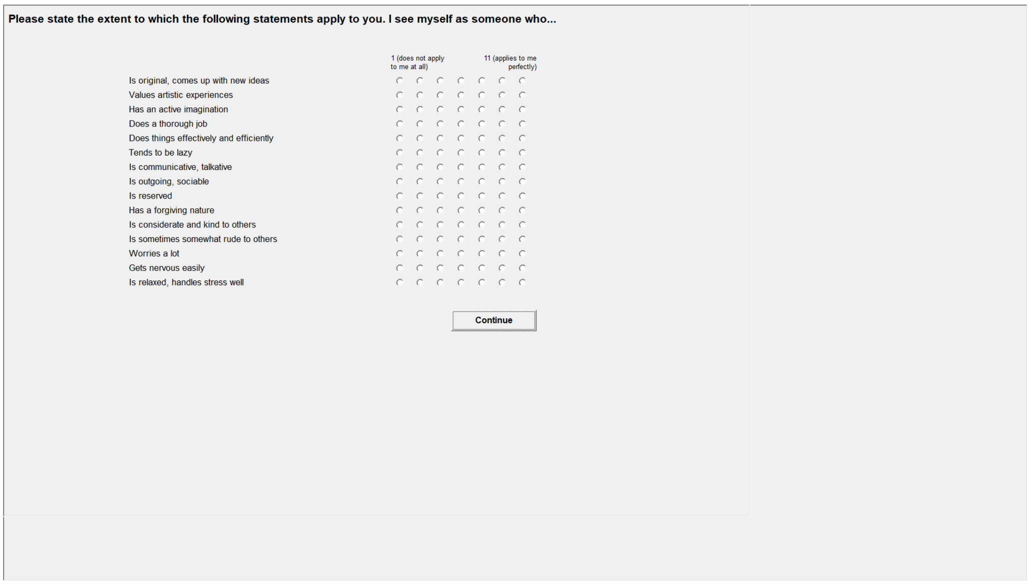


Figure 86: Page 28

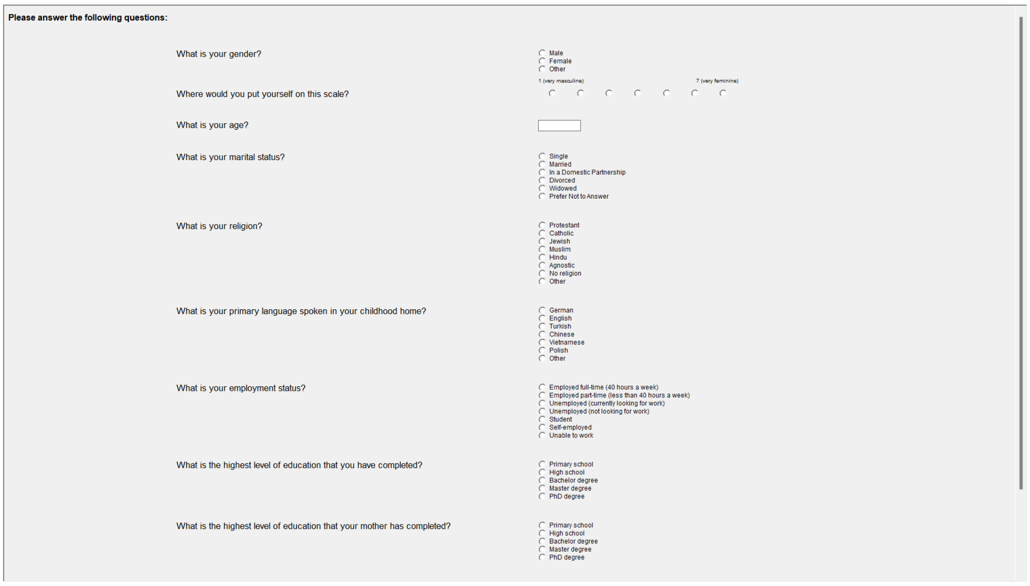
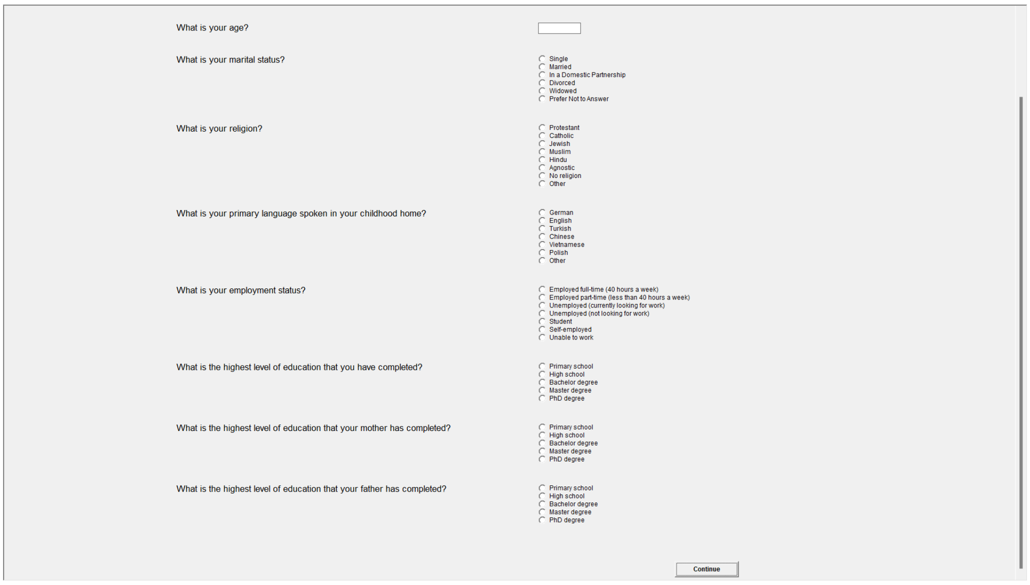


Figure 87: Page 29


What is your age?

What is your marital status?
- Single
- Married
- In a Domestic Partnership
- Divorced
- Widowed
- Prefer Not to Answer

What is your religion?
- Protestant
- Catholic
- Jewish
- Muslim
- Hindu
- Agnostic
- No religion
- Other

What is your primary language spoken in your childhood home?
- German
- English
- Turkish
- Chinese
- Vietnamese
- Polish
- Other

What is your employment status?
- Employed full-time (40 hours a week)
- Employed part-time (less than 40 hours a week)
- Unemployed (currently looking for work)
- Unemployed (not looking for work)
- Student
- Self-employed
- Unable to work

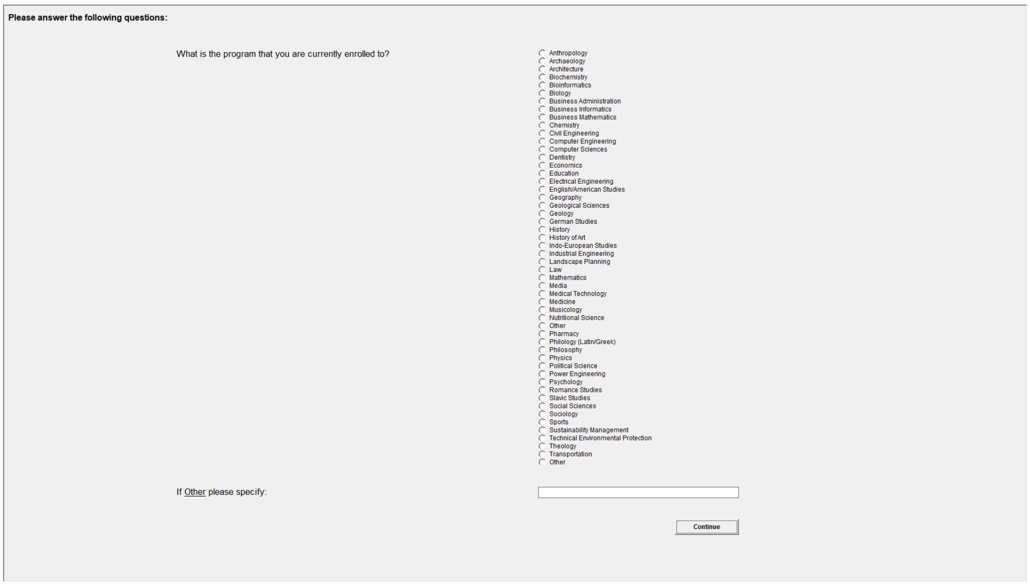
What is the highest level of education that you have completed?
- Primary school
- High school
- Bachelor degree
- Master degree
- PhD degree

What is the highest level of education that your mother has completed?
- Primary school
- High school
- Bachelor degree
- Master degree
- PhD degree

What is the highest level of education that your father has completed?
- Primary school
- High school
- Bachelor degree
- Master degree
- PhD degree

Continue

Figure 88: Page 30


Please answer the following questions:

What is the program that you are currently enrolled to?
- Anthropology
- Archaeology
- Architecture
- Biochemistry
- Bioinformatics
- Biology
- Business Administration
- Business Informatics
- Business Mathematics
- Chemistry
- Civil Engineering
- Computer Engineering
- Computer Sciences
- Dentistry
- Economics
- Education
- Electrical Engineering
- English/American Studies
- Geography
- Geological Sciences
- Geology
- German Studies
- History
- History of Art
- Indo-European Studies
- Industrial Engineering
- Landscape Planning
- Law
- Mathematics
- Media
- Medical Technology
- Medicine
- Musicology
- Nutritional Science
- Other
- Pharmacy
- Philology (Latin/Greek)
- Philosophy
- Physics
- Political Science
- Power Engineering
- Psychology
- Romance Studies
- Slavic Studies
- Social Sciences
- Sociology
- Sports
- Sustainability Management
- Technical Environmental Protection
- Theology
- Transportation
- Other

If Other please specify:

Continue

Figure 89: Page 31

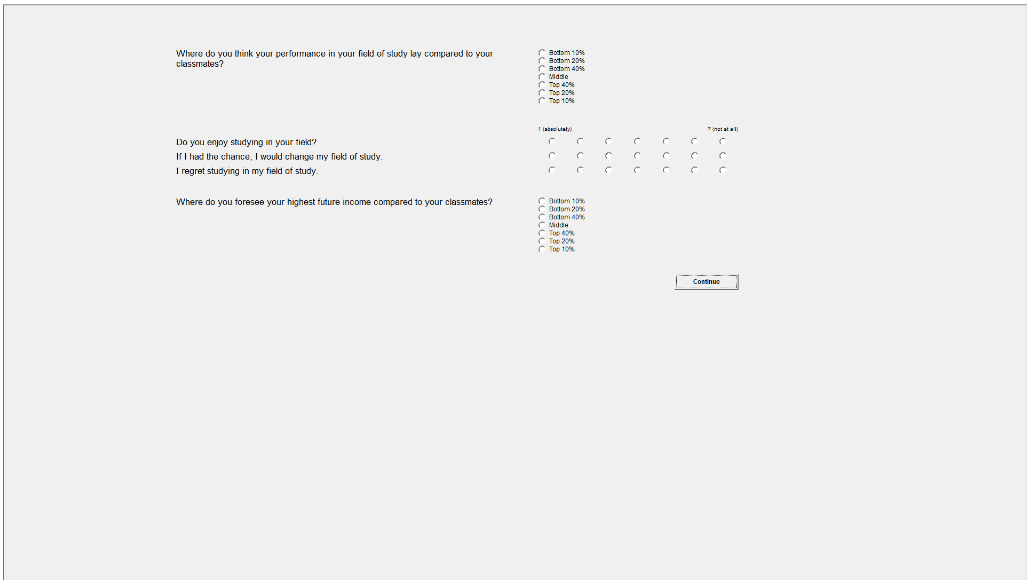


Figure 90: Page 32

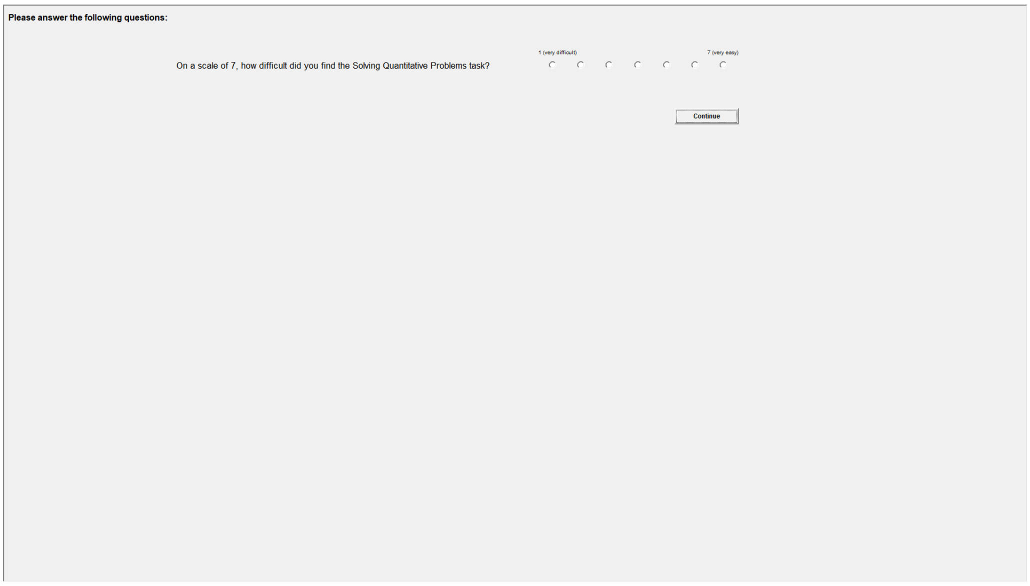


Figure 91: Page 33

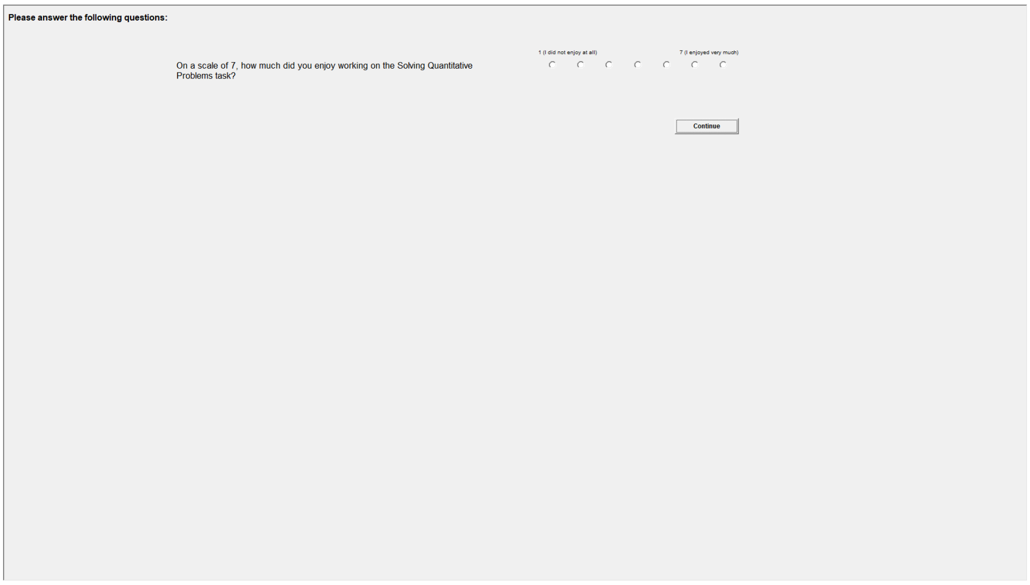


Figure 92: Page 34

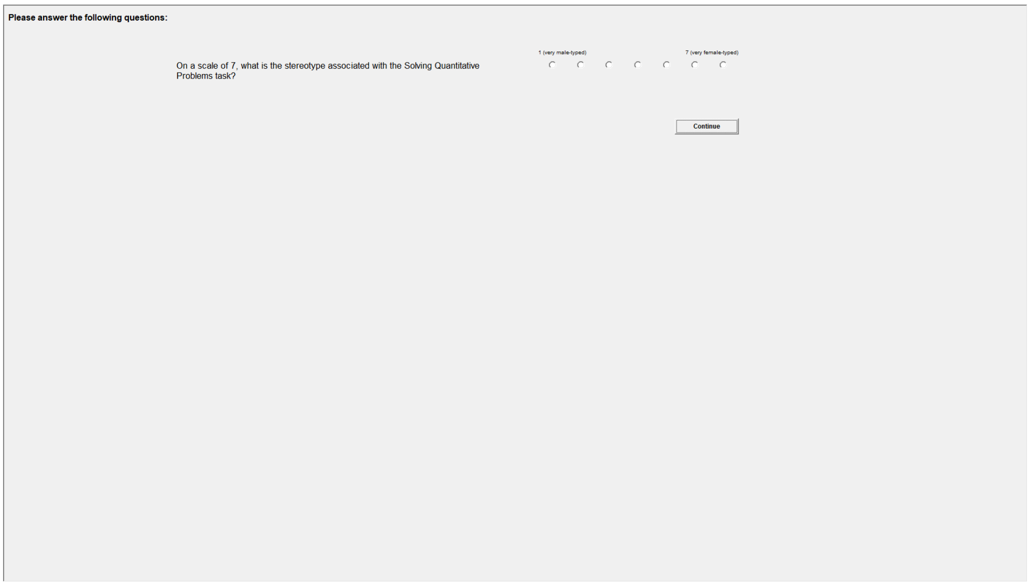


Figure 93: Page 35

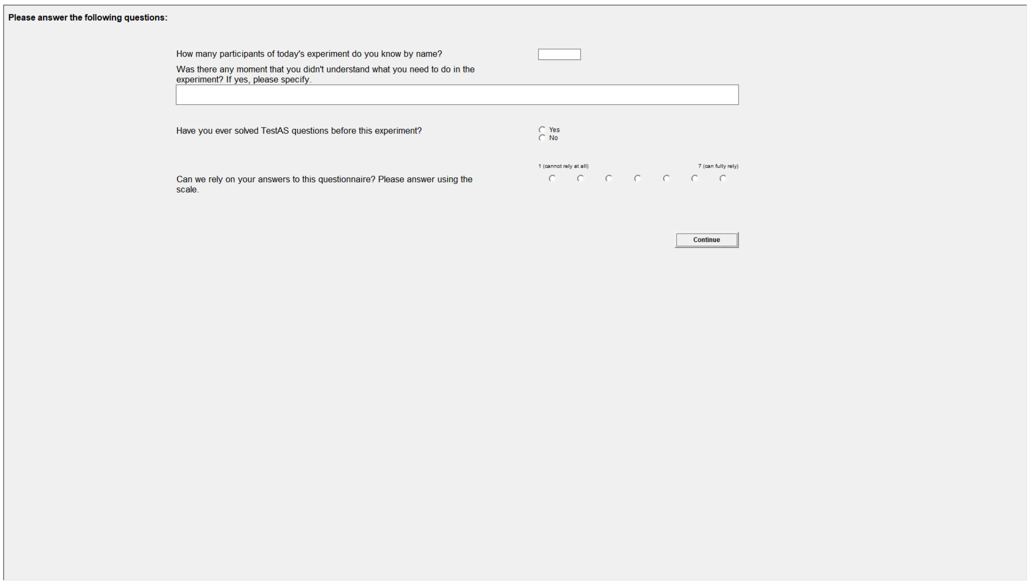


Figure 94: Page 36