Pushing the Limit: The Different Effects of Threshold Proximity and Competitiveness

Arnaud Chevalier* Francesco Feri† Parvaneh Pasandi† Enya Turrini†

Abstract

This study examines how performance thresholds interact with competitiveness and gender to influence effort and goal achievement. In a laboratory experiment, participants complete a baseline task, reveal their willingness to compete through an incentive-choice decision, and then face a personalized performance target of varying difficulty. The results show clear gender differences: men are more likely to choose competitive incentives, respond more strongly to the introduction of a goal, and achieve targets at higher rates than women, despite similar baseline performance. More attainable goals independently increase effort and success for all participants. While competitiveness is positively related to goal achievement, mediation analysis indicates that it does not account for the gender gap in outcomes. Robustness checks suggest that individual traits, such as self-confidence, play an additional role. Overall, the findings demonstrate that performance thresholds are not neutral incentives and that their effectiveness depends on both goal design and individual characteristics.

Keywords: Goal setting, Competitiveness, Gender differences

JEL Classification: C91, D91, J16, M52

1 Introduction

Many people set goals, but not everyone manages to reach them. Some individuals push through and succeed, while others fall just short. This pattern can be observed in the field: salespersons aiming to hit a sale target, students trying to get a higher grade, or runners chasing a personal best. However, whether they can hit their goal by increasing their effort, depends on different factors. This study investigates why such differences arise, focusing on how goal proximity, competitiveness, and gender influence effort and the likelihood of success in reaching that goal. Specifically, we ask (i) whether being closer to a threshold motivates people to exert more effort, (ii) how individual competitiveness shapes behavior, and (iii) whether men and women respond differently to these challenges.

By answering these questions, we aim to better understand the drivers of persistence and improvement when individuals face performance targets. The application of appropriate mechanisms to push people towards reaching a higher individual goal could have

We are grateful to the faculty and colleagues in the Department of Economics at Royal Holloway, as well as to the participants of BEE UK 2025, for their valuable comments.

*Royal Holloway, University of London and IZA

†Royal Holloway, University of London

various implications: workers could improve the productivity of the firm depending on how their production targets are set, or students could achieve higher grades, however, people may respond differently to the same goal due to willingness to compete (with others and with themselves) and gender differences. Hence, setting personalised thresholds based on such qualities could lead to everyone putting more effort, at work or at school, implying higher personal - and even collective - gains.

Existing research shows that both goal-setting and competitive environments shape individual effort, yet the interaction between these two dimensions remains less understood. While studies have shown that goals or thresholds can motivate individuals to increase performance (Locke & Latham, 2002), and a large body of work documents gender differences in competitive behaviour (Niederle & Vesterlund, 2007), it is still unclear how competitiveness influences responses to personalised performance targets. Moreover, we know little about whether the gender patterns observed in competitive settings also emerge when individuals face thresholds tailored to their own ability. Our study contributes to filling this gap by examining how goal proximity, competitiveness, and gender jointly affect effort when individuals are presented with personalised thresholds.

To investigate these questions, we conducted a controlled laboratory experiment centred on a letter-counting task. Participants completed the same task across four stages, each featuring a distinct payoff scheme. The first stage provides a baseline measure of effort. The second stage introduces a tournament in which participants compete in groups of four. In the third stage, participants choose between the tournament and a piece-rate scheme, following the Niederle and Vesterlund (2007) framework for measuring competitiveness. In the final stage, participants face an incentivized performance threshold, personalised based on their baseline performance and treatment assignment.

The experiment also incorporates additional measures to provide context for individual differences. Competitiveness is assessed through revealed choices in incentive schemes, complemented by self-reported measures, while gender, risk attitudes, and belief-based measures are also collected. Taken together, these features allow us to identify which factors best explain why some individuals rise to the challenge of performance thresholds, while others fall behind.

Our findings reveal substantial heterogeneity in responses to performance goals. Pronounced gender differences emerge throughout the experiment: men are more likely to select competitive incentives, respond more strongly to the introduction of performance targets, and achieve these targets at higher rates than women, despite comparable baseline performance. Goal proximity also plays a crucial role. Participants facing more attainable thresholds increase their effort more and are more likely to succeed than those confronting more demanding goals. While competitiveness is positively associated with goal achievement, mediation analysis shows that it does not account for the observed gender gap in outcomes, suggesting that additional mechanisms beyond competitive preferences shape performance under threshold incentives.

The remainder of this chapter is structured as follows. Section 2 reviews the related literature on goal-setting, competitiveness, and gender differences in performance. Section 3 introduces a theoretical model of effort provision under performance thresholds. Section 4 describes the experimental design and procedures, and Section 5 presents the hypotheses. Section 6 reports the experimental results, including robustness analyses. Section 7 concludes.

2 Related Literature

Our paper contributes to two strands of literature: (i) the literature on behaviour under threshold or goal provision, and (ii) the literature on gender differences in competitiveness and performance.

A growing literature in experimental economics examines how individuals respond to the provision of goals or thresholds. Corgnet et al. (2015) show that setting attainable yet challenging goals can significantly increase workers' performance, even when these goals are purely intrinsic and not tied to any explicit rewards or punishments, implying that goal-setting itself can act as a behavioural motivator. Similarly, Abeler et al. (2011) demonstrate that individuals adjust their effort in response to expectations: when expected performance levels are high, subjects work longer hours and ultimately earn more compared to situations with lower expectations. Together, these studies suggest that thresholds or expectations can influence effort provision even in the absence of direct monetary incentives.

Focusing instead on gender and performance in competitive settings, Gneezy et al. (2003) document that men tend to perform better in competitive environments, and that women's performance decreases particularly when competing against men rather than against other women. In contrast, Niederle and Vesterlund (2007), our main paper of reference for the design, find no gender gap in performance but identify a substantial gender difference in willingness to enter competitive environments: men choose to compete significantly more often than women. Markowsky and Beblo (2022) in a meta-analysis showed that larger gender differences in willingness to compete are observed in studies that are more similar to the Niederle and Vesterlund (2007) setting, i.e., student pools with math tasks, but almost negligible for other age groups and different tasks. Subsequent papers report more nuanced findings. On the other hand, De Paola et al. (2015) find no gender differences either in willingness to compete or in performance under competition. Analysing the role of confidence in willingness to compete, Lénárd et al. (2024) show that gender gaps in competitiveness appear only for individuals with realistic confidence, but not for those who are over- or under-confident.

Despite extensive work on competitiveness and goal-setting, little is known about how competitiveness interacts with personalised thresholds or individually tailored goals. Moreover, it remains unclear whether gender differences observed in competitive environments also emerge when individuals face threshold-based incentives or personalised performance targets. Our paper addresses these gaps by investigating how gender and competitiveness shape responses to personalised thresholds in an experimental setting.

Competition with others and proneness to reach a tailored threshold are indeed likely related, as we can think of a goal-setting environment as competing against one-self. Apicella et al. (2020) show that allowing individuals to compete against their own past performance substantially increases the share of participants who choose to compete, with both men and women strongly preferring self-competition to competing against others. Hence, while women especially may shy away from competing against others, this gender gap reduces when self-competition is involved.

3 The Theoretical Framework

This section develops a two-stage model of effort provision in a task with a performance threshold defined in terms of completion time. The model mirrors the experimental design

in which individuals first complete a task under baseline incentives and then face an additional prize if they improve their performance relative to their own baseline. Individuals are not aware of the performance threshold of Stage 2 when they are performing Stage 1. The framework highlights how the distance to a threshold affects both the likelihood and the magnitude of effort adjustments.

3.1 Setup

Time is divided into two stages, indexed by $t \in \{1, 2\}$. In each stage, the individual chooses an effort level $e_t \in [0, 1]$, where higher effort reduces the time required to complete the task but entails a personal cost. The payoff structure in each stage is revealed to the individual only at the start of that stage.

Let $T > 0$ denote the maximum time available to complete the task. Completion time depends linearly on effort according to:

$$t(e) = T(1 - e). \tag{1}$$

Thus, exerting zero effort implies that the task takes the full available time T, while higher effort shortens completion time.

Effort is costly. Preferences in each stage are represented by the following payoff function:

$$\pi(e) = T - t(e) - \frac{c}{2}e^2, \tag{2}$$

where $c > T$ captures the marginal cost of effort. Substituting for $t(e)$ yields:

$$\pi(e) = Te - \frac{c}{2}e^2. \tag{3}$$

Stage 1: Baseline Performance

In Stage 1, the individual chooses effort e_1 to maximize:

$$\max_{e_1 \in [0,1]} \; Te_1 - \frac{c}{2}e_1^2. \tag{4}$$

The first-order condition is:

$$T - ce_1 = 0, \tag{5}$$

which yields the optimal baseline effort:

$$e_1^* = \frac{T}{c}. \tag{6}$$

The corresponding baseline completion time is:

$$t_1^* = T(1 - e_1^*) = T\left(1 - \frac{T}{c}\right), \tag{7}$$

and the associated payoff is:

$$\pi_1^* = \frac{T^2}{2c}. \tag{8}$$

Baseline performance in Stage 1 serves as the reference point for the performance threshold in Stage 2. Note that the optimal effort in Stage 1 is chosen without anticipation of the threshold or incentives introduced in Stage 2, as this information is revealed to the individual only at the beginning of Stage 2.

Stage 2: Threshold Incentives

In Stage 2, the individual again chooses effort e_2, but now faces a performance threshold defined in terms of completion time. Let θ denote the threshold time. The threshold is defined as a proportion of the individual's baseline completion time:

$$\theta = \alpha t_1^*, \tag{9}$$

where $\alpha \in (0, 1)$. Higher values of α correspond to less demanding thresholds.

If the individual completes the task in time θ or faster, they receive an additional fixed prize $P > 0$.

The threshold expressed in terms of effort is obtained from the condition $t(e_2) \leq \theta$. Substituting the time function into this condition yields:

$$T(1 - e_2) \leq \alpha T(1 - \frac{T}{c})$$

$$e_2 \geq \bar{e}(\alpha) \equiv 1 - \alpha \left(1 - \frac{T}{c}\right) \equiv 1 - \alpha \left(1 - e_1^*\right). \tag{10}$$

Payoffs in Stage 2 are therefore given by:

$$\pi_2(e_2) = \begin{cases} Te_2 - \frac{c}{2}e_2^2, & \text{if } e_2 < \bar{e}(\alpha), \\ Te_2 - \frac{c}{2}e_2^2 + P, & \text{if } e_2 \geq \bar{e}(\alpha). \end{cases} \tag{11}$$

3.2 Results

Optimal Effort Conditional on Meeting the Threshold

Proposition 1. If the individual chooses to meet the threshold in Stage 2, the optimal effort choice is $e_2 = \bar{e}(\alpha)$.

Proof. For any $e_2 \geq \bar{e}(\alpha)$, the prize P is fixed. The marginal benefit of effort is therefore T, while the marginal cost is ce_2, which is strictly increasing in effort. The unconstrained optimum satisfies $e = T/c = e_1^* < \bar{e}(\alpha)$. Hence, for all $e_2 \geq \bar{e}(\alpha)$, marginal costs exceed marginal benefits, and any effort above the threshold reduces payoff. Therefore, conditional on meeting the threshold, the optimal choice is $e_2 = \bar{e}(\alpha)$.

Incentives to Increase Effort

In Stage 2, the individual faces a discrete decision of whether to increase effort relative to baseline performance. Maintaining baseline effort yields the same payoff as in Stage 1, while increasing effort requires raising effort from e_1^* to at least the threshold level $\bar{e}(\alpha)$ and is therefore costly. The individual increases effort if and only if the fixed prize P compensates for this additional effort cost.

The payoff from maintaining baseline effort is:

$$\pi_2^B = Te_1^* - \frac{c}{2}(e_1^*)^2. \tag{12}$$

The payoff from increasing effort to meet the threshold is:

$$\pi_2^\theta = T\bar{e}(\alpha) - \frac{c}{2}\bar{e}(\alpha)^2 + P. \tag{13}$$

Increasing effort is optimal if and only if $\pi_2^\theta \geq \pi_2^B$ (i.e. $13 \geq 12$) that implies:

$$P \geq \underline{P}(\alpha) \equiv \frac{c}{2}\left[\bar{e}(\alpha)^2 - (e_1^*)^2\right] - T\left[\bar{e}(\alpha) - e_1^*\right]. \tag{14}$$

Proposition 2. For a fixed prize P, there exists a cutoff value α^* such that the individual increases effort in Stage 2 if and only if $\alpha \geq \alpha^*$.

Proof. The effort threshold $\bar{e}(\alpha) = 1 - \alpha(1 - T/c)$ is decreasing in α, implying that higher values of α require smaller effort increases relative to baseline. Differentiating $\underline{P}(\alpha)$ with respect to α yields:

$$\frac{d\underline{P}(\alpha)}{d\alpha} < 0. \tag{15}$$

Therefore, for a fixed prize P, the condition $P \geq \underline{P}(\alpha)$ holds for higher values of α, implying that for an individual it is optimal to meet the threshold only when the threshold is sufficiently close to baseline performance.

The model predicts that baseline effort is uniquely determined by the trade-off between time savings and effort costs. In the presence of a performance threshold, individuals face a discrete decision of whether to increase effort at all. Conditional on responding, they optimally choose the minimum effort necessary to meet the threshold.

These predictions provide a theoretical foundation for the experimental design. In particular, the framework highlights how varying the difficulty of the threshold influences the decision to exert additional effort among our treatments.

4 Experiment Design

Players take part in an effort task where working at a faster pace leads to higher payoffs. The same task is repeated across four stages, each with different payoff mechanisms to capture different behaviors. The settings and purpose of each stage are explained below.

The task is a letter-counting activity: players see a grid of random letters and are asked to count how many times a target letter appears. After submitting their answer, a new grid with a new target letter is shown. To finish the task, players must reach a set number of correct answers (5), and the quicker they do so, the higher their potential payoff will be. (See Figure 1)

The total time available to complete the task is limited to 5 minutes. While we expect most players to finish within this limit, some may run out of time before submitting all the required correct answers. We refer to these participants as *timeout* players. Another group who may fail to finish are those who reach the maximum number of wrong submissions. To prevent random guessing, we set a strict limit on incorrect answers (5), after which the task stops and the player receives no payoff. These participants are referred to as *max-wrong* players. Participants who failed to submit at least one correct answer within the 5-minute time limit in Stage 1 were immediately directed to an exit page and excluded from the study.[1]

To allow participants to familiarise themselves with the task, they first complete a 3-minute practice round, which has no impact on either their payoff or their potential exclusion from the study.

This task was chosen following Abeler et al. (2011) and Charness et al. (2018). In their study on the relationship between reference points and the provision of effort, the authors

[1] In our sample, no participants were excluded based on this criterion.

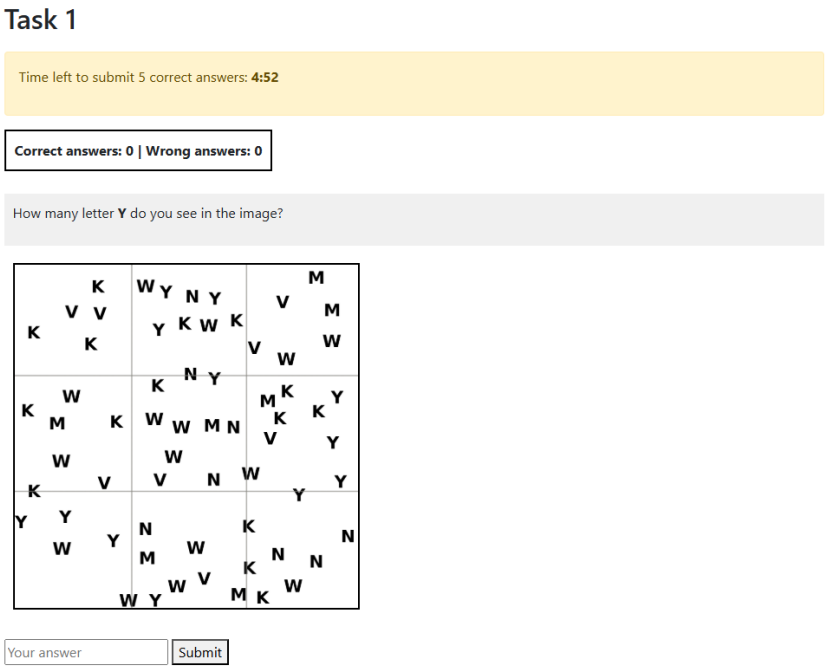

Figure 1: Screen of Counting Letter Task

employ a "counting zeros" task, which does not require prior knowledge and for which learning effects are minimal. They also report no gender differences in performance on this task. Given our interest in gender-related aspects in performance and in competition, selecting a gender-neutral task rather than one perceived as male- or female-oriented is of particular importance. Additionally, the potential for participants to cheat or use AI tools to complete the task is minimised. Since our study was conducted online through Prolific (Palan & Schitter, 2018), we could not employ the mathematical task used in Niederle and Vesterlund (2007). Instead, we opted for a task that reduces the incentive to rely on AI assistance. Using AI would likely result in time loss and an increased likelihood of submitting incorrect answers. [2]

The following section explains how payoffs are determined in each stage. Participants receive the relevant instructions for each stage only at the beginning of that stage. This avoids any strategic anticipations. All participants completed the four stages in the same order. For payment purposes, one of these stages was randomly selected at the end of the experiment.

4.1 Stage 1: Individual Piece-Rate Task

In the first stage, players complete the letter-counting task individually without any threshold requirement. They have 5 minutes to submit 5 correct answers, with a maximum of 5 wrong attempts allowed. Payoffs are calculated as the number of correct answers (up to 5) plus the number of seconds left on the timer:

$$\text{Piece-Rate Payoff} = (\text{time remaining} + 1 \times \text{correct answers})$$

[2]We acknowledge that a letter-counting task with imperfectly aligned letters could pose difficulties for participants with dyslexia. Nonetheless, prior to participation, subjects were informed that the task might not be suitable for individuals with dyslexia.

For example, a player who finishes in 4 minutes earns 60 points for the remaining 60 seconds and 5 points for the 5 correct answers. A *timeout* player who manages only 4 correct answers before the time runs out receives 4 points for accuracy but no time bonus, since no seconds are left. A *max-wrong* player who reaches the limit of wrong submissions before completing 5 correct answers receives zero points (and is excluded from the study if there are no correct answers submitted).

This stage is used to measure each player's baseline performance, which we define as the time at which the player submits their last correct answer. Using the total time spent to finish the task would not give us a fair baseline for timeout or max-wrong players. For example, all timeout players would appear to have a baseline of zero, even though one might have been only seconds away from completing the fifth correct answer, while another managed only a single correct answer before running out of time. These two cases are very different, yet both would be treated as the same under a zero baseline. This is not a fair baseline for max-wrong players either, since the task ends immediately after their fifth wrong submission, which is no indication of the time spent to correctly complete the task. To ensure comparability and to include as many players as possible, we therefore set the baseline as the time of the last correct submission, taking into account both the timing and the number of correct answers achieved.

After finishing, players receive feedback showing their number of correct and incorrect answers, along with the time they spent.

4.2 Stage 2: Tournament

In Stage 2, each player competes in a tournament against 3 other players. This helps us observe performances under competition, and it is also necessary for Stage 3 where we measure their willingness to compete (Niederle & Vesterlund, 2007).
Players are randomly assigned into groups of four, and the fastest player in each group becomes the winner. The tournament follows a winner-takes-all payoff rule. The winning player in each group earns four times the piece-rate payoff from Stage 1, that is:

$$\text{Winner's Payoff} = (\text{time remaining} + 1 \times \text{correct answers}) \times 4$$

Which will be received if this stage is randomly selected for payment. All other players in the group receive zero. This payment structure is taken and adapted from Niederle and Vesterlund (2007), designed so that the chances to win for a given performance is 0.25 and the winner's payoff is four times the individual payoff from Stage 1. This ensures that the expected value of the tournament in Stage 2 is equal to that of the individual scheme in Stage 1.

The task is the same letter-counting task where they have 5 minutes to solve 5 grids correctly, with a maximum of 5 wrong submissions allowed.[3]

To determine the winner, we rank players in two steps. First, players are ordered by the number of correct answers submitted, from highest to lowest. Second, if two or more players have the same number of correct answers, they are ranked by the time of their last correct submission, with faster completions ranked higher. This system allows us to compare even players who end the task with a timeout or by reaching the maximum number of wrong answers.

[3]However, if at this stage they do not submit any correct answer, they are still allowed to continue the study.

The ranking results and the winner are not revealed to the player to avoid affecting their decision and performance in the following stages.

4.3 Stage 3: Choice of payoff scheme

At this stage, players have already experienced the task under both the individual and tournament payoff schemes. They are now asked to choose their preferred scheme and complete the task once again. This gives us an indicator of the willingness to compete and whether some players shy away from competing. For a given performance level, the payoffs are equal in expectations between the piece-rate (Stage 1) and the tournament (Stage 2), hence choosing the tournament may signal confidence in one's relative performance, a general willingness to compete, or attitudes toward risk. To separate competitiveness from risk preferences, we later measure players' risk attitudes through a self-report survey included at the end of the experiment.

Players who choose the tournament are evaluated against their group members' Stage 2 performances, not against others' Stage 3 choices. This design ensures that everyone is compared under the same compensation scheme, avoids biases from beliefs about others' choices, and removes externalities where a win could harm others' payoffs. As a result, the decision in Stage 3 is effectively an individual choice, based only on a player's ability to outperform past tournament results and their preference for competition.

4.4 Stage 4: Threshold

In this stage, we set a personalized threshold for the players to incentivize them to improve their initial performance. The goal is to submit the same number of correct answers that the player was able to submit in Stage 1,[4] however, within a lower time compared to Stage 1, hence applying a time threshold.

The time threshold is lower than the time when the last correct answer was submitted in Stage 1. The extent to which the time is decreased depends on the treatments. We have two treatments in this Stage that players get randomly assigned to. In the *Near* treatment, the time threshold is 5% lower, and in the *Far* treatment, it is 10% lower than the last correct submission time of Stage 1.

Payoffs in this stage are calculated in the same way as in the individual payoff scheme of Stage 1, with an additional 20 points awarded if the player meets their personalized goal. Hence, to receive the bonus payment, players need to submit as many correct answers as they did in Stage 1, and do that within the time threshold applied. If one of these conditions is not met, participants are paid the normal piece-rate payoff, with no bonus. All payoffs are subject to whether that stage is the randomly selected stage for payment at the end of the study.

Players are not aware of the personalized threshold while performing Stage 1. This ensures that they cannot act strategically to secure an easier threshold later.

4.5 Post Experimental Questions

Throughout the experiment, we elicit participants' beliefs about their own performance using three incentivized guessing questions administered at different stages. After Stage

[4]This allows us to include those players who were not able to complete the task with 5 required correct submissions.

1, participants are asked to assess their relative performance compared to the average participant by indicating whether they believe they performed above, below, or about the same as the average (GQ1). After Stage 2, participants are asked to state their best guess about their rank within their four-member tournament group (GQ2). Finally, prior to Stage 4, after the participants are informed about their personalized threshold, they report whether they believe they will meet the performance threshold (GQ3).

All belief elicitation questions are incentivized. Participants earn an additional 10 points if their stated belief corresponds to their realized performance; however, to limit hedging across questions, only one of the three guessing questions is randomly selected for payment. Importantly, participants receive no feedback on the accuracy of their guesses until the end of the experiment, ensuring that beliefs are not updated through interim feedback.

The study is also followed by a questionnaire where demographic data are collected, as well as Big Five (BFI-10, Soto and John (2017)), and risk attitude (Falk et al., 2023). In order to further assess their attitude towards competition, we include two questions following Fallucchi et al. (2020). While their initial version of the question is phrased as "Competition brings the best out of me", we extend it as "Competition brings the best out of me, *especially if I'm competing against others*", and "Competition brings the best out of me, *especially if I'm competing against myself, without other competitors involved*", both to be answered on a 7-point likert scale.

Throughout the experiment, multiple control questions are asked to account for participant's attention and understanding of the instructions.

The full set of instructions is provided in Appendix **??**.

4.6 Experimental Procedure

Ethical approval was obtained from Royal Holloway University of London, and participants could advance in the study only upon agreement to the consent form. The experiment was preregistered on AsPredicted (#244198).

Participants were recruited via Prolific (Palan & Schitter, 2018) on the 9^{th} of October 2025, with respect to the UK population and quotas by gender. The study was programmed in OTree (Chen et al., 2016), and participants could decide to abandon the study at any time (implying though no participation fee being paid). The median time for the study was 25 minutes, and the average payment of £15.28, inclusive of a participation fee of £6.00 (aligned with Prolific policies and similar studies, see e.g., Peer et al. (2022), and Albert and Smilek (2023)).

4.7 Sample

The sample consists of 100 subjects, equally divided between the two treatments. With respect to gender, we have 26 (24) females (males) in the *Near* treatment, and instead 24 (26) females (males) in the *Far* Treatment.

Table 1 shows the descriptive statistics by gender, specifically a balance table following Chiapello (2018). The sample is balanced by gender, with statistically significant differences at 5% for age and economic activity.[5] Looking into the control questions,

[5]As reported in the OSM, there are no statistically significant differences among the demographic characteristics between treatments.

no subjects answered to all of them wrong, however, 84% (92%) of males (females) answered at least one control question wrong. Further descriptive statistics (balance table by treatments, full descriptive statistics, details on control questions) can be found in the OSM.

Table 1: Descriptive Statistics by Gender

Variable	(1) Male	(2) Female	(3) Difference
Age	40.040	46.520	6.480*
	(12.015)	(14.009)	(2.610)
Education level	5.620	5.740	0.120
	(0.901)	(0.965)	(0.187)
Economic activity	1.880	2.920	1.040*
	(1.649)	(2.481)	(0.421)
Study/work field	6.300	6.360	0.060
	(1.992)	(2.593)	(0.462)
Control Questions - at least 1 wrong (=1)	0.840	0.920	0.080
	(0.370)	(0.274)	(0.065)
Control Questions - all wrong (=1)	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Observations	50	50	100

Notes: Standard deviations in parentheses. Stars indicate significance of mean differences: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

4.8 Hypotheses

Guided by prior evidence on gender differences in competitiveness and performance under incentives, and based on our theoretical model, we formulate five directional hypotheses.

H1: Men are more likely to select into competition than women.

H2: Conditional on the presence of a performance threshold, we expect performance improvements to be larger when the threshold is near rather than far.

H3: Men are more likely than women to achieve the goal defined by the threshold.

H4: Introduction of a threshold induces larger performance improvements for men than for women.

H5: Competitive individuals are more likely to meet their personalized threshold.

All hypotheses are pre-registered (AsPredicted #244,198) except for the last hypothesis (H5).

5 Results

This section presents the main empirical findings of the study. We begin, in the next subsection, by examining competitiveness, as inferred from participants' performance and payoff-scheme choices in Stages 1 through 3. We then turn to Stage 4 to analyze goal achievement and performance responses under the threshold incentive. The Econometrics subsection introduces a structural model that estimates whether competitiveness mediates the relationship between gender and goal achievement. Finally, the Robustness Checks employ a two-stage residual inclusion approach to assess whether competitiveness is endogenous in the goal-achievement equation and also re-estimate the structural model controlling for additional variables.

First, we define some of the terms used in this section before starting the analysis. Throughout the paper, we define *performance* as the time taken to complete the task. Task completion is achieved when a participant submits five correct answers before the timer expires. To ensure that faster completion does not come at the expense of accuracy, we first examine variations in accuracy across stages. As the accuracy of participants remains consistent throughout the stages (see Table 2), this suggests that improvements in speed are not driven by reduced accuracy. Therefore, we can confidently use speed as the primary measure of performance.

Table 2: Count of correct answers per stage - accuracy remains consistent

Correct answers	Stage 1	Stage 2	Stage 3	Stage 4
1	0	1	1	3
2	2	0	0	0
3	1	3	2	1
4	4	3	4	4
5	93	93	93	92
Total	100	100	100	100

Self-confidence is measured using participants' beliefs about their relative performance. This approach follows the seminal design of Niederle and Vesterlund (2007), who elicit subjects' guessed rank as a proxy for confidence, and has since become standard in the experimental literature on competitiveness and gender differences. A large body of subsequent work similarly relies on incentivized belief elicitation about own performance to capture confidence or self-assessment, as summarized in recent reviews and meta-analyses (e.g. Horn et al. (2022); Markowsky and Beblo (2022); Van Veldhuizen (2022)).

An alternative strand of the literature argues that beliefs should be evaluated relative to actual performance, distinguishing overconfidence, underconfidence, and accurate self-assessment based on the gap between perceived and realized outcomes (e.g. Lénárd et al. (2024)). While informative for isolating miscalibration, such measures conflate confidence with performance accuracy and may obscure the role of beliefs themselves in driving economic decisions. In the context of tournament entry, what matters is not whether beliefs are correct ex post, but whether individuals believe they are strong relative to others when making their choice.

Accordingly, we adopt the Niederle–Vesterlund approach and measure self-confidence based on participants' stated beliefs about their performance, independently of realized

outcomes. Beliefs are elicited through three incentivized guessing questions administered at different stages of the experiment, allowing us to construct a composite measure of self-confidence that captures persistent differences in perceived relative ability.

To construct a unified measure of self-confidence, we assign numerical scores to each belief response. For GQ1, a response of *above average* is coded as +1, *about the same* as 0, and *below average* as −1. For GQ2, reporting rank 1 is coded as +1, rank 4 as −1, and ranks 2 or 3 as 0. For GQ3, believing that one will meet the threshold is coded as +1, while believing that one will not meet the threshold is coded as −1. The three scores are then summed to obtain an aggregate self-confidence index ranging from −3 to +3, with higher values indicating stronger beliefs in one's own relative ability and likelihood of success.

This composite measure captures confidence across multiple performance contexts within the experiment and reflects persistent differences in perceived ability rather than stage-specific beliefs. A summary of the scoring procedure is provided in Table 3. For each question, responses are mapped to scores of −1, 0, or +1, reflecting pessimistic, neutral, or optimistic beliefs, respectively. The table also reports the proportion of participants selecting each response, illustrating substantial variation in beliefs across all three elicitation points.

Table 3: Construction of the Self-Confidence Index

Score	GQ1: Relative Performance	GQ2: Rank in Group	GQ3: Threshold
+1	Above average (31%)	Rank 1 (23%)	Yes (56%)
0	Similar to avg (52%)	Rank 2 or 3 (70%)	–
−1	Below average (17%)	Rank 4 (7%)	No (44%)

Notes: The self-confidence index is the sum of the three scores and ranges from −3 to +3. In parantheses are the actual response rates in percentages. N = 100.

Risk attitudes are measured using a self-reported willingness-to-take-risks question on a 0–10 scale, following Falk et al. (2023). Table 4 reports descriptive statistics by gender. Men report higher risk tolerance than women, with an average score of 6 compared to 4.08 for women. While these measures are self-reported and not incentivized, they indicate meaningful heterogeneity in risk preferences across participants, with men exhibiting greater stated willingness to take risks in this sample.

Table 4: Risk Attitudes by Gender

Gender	Mean	SD	Median (p50)	*N*
Male	6	2.33	7	50
Female	4.08	2.55	3	50
Total	5.04	2.62	5.5	100

Risk attitudes asked through a self-report question following Falk et al. (2023) on a scale from 0 to 10 where 0 is *not willing to take risks at all*.

5.1 Competitiveness and Gender Effects

Competitiveness is measured through the revealed choice in Stage 3, where players decide whether to be paid under the individual piece-rate or the tournament scheme. Under a given performance level, where everyone has equal chances to win, the expected payoff from the tournament is equal to the expected piece-rate payoff. The tournament payoff is four times the piece-rate payoff, and the probability of winning in a group of four is 0.25. According to this setting, a payoff-maximizing player should be indifferent between the two schemes. Therefore, choosing the tournament reflects a willingness to compete rather than maximizing expected monetary advantage.

Unlike Niederle and Vesterlund (2007), who report no gender differences in performance under either piece rate or tournament payoffs, we observe some gender differences in performance, particularly in Stage 1 (see Figure 2). Women complete the task in Stage 1 approximately 17 seconds faster than men, and this difference is statistically significant based on a paired t-test ($p = 0.04$). Under the tournament scheme, men are on average slightly faster than women, although this difference is not statistically significant.

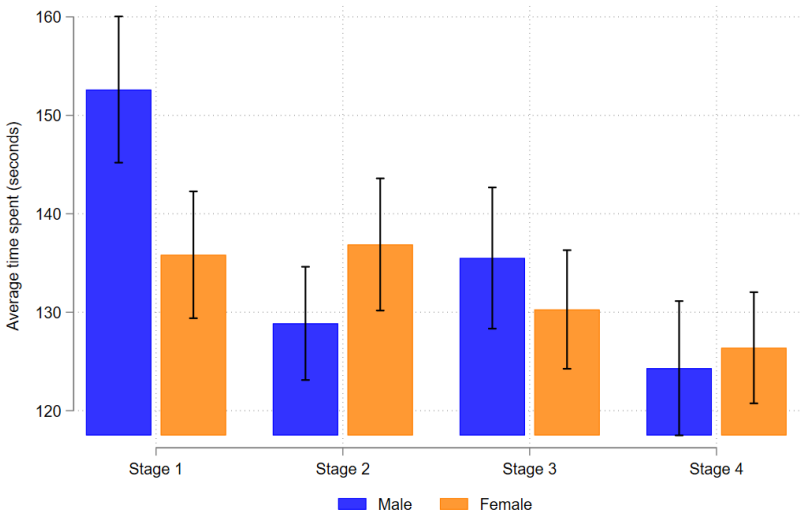

Figure 2: Average time spent across stages by gender. Bars are standard errors.

Although women perform faster in Stage 1, these performance differences do not explain tournament entry. Table 5 shows the results of a logistic regression on the choice of tournament entry. When we regress the tournament choice on baseline performance (Stage 1 time), performance under the tournament (Stage 2 time), and a gender indicator, neither Stage 1 performance nor Stage 2 performance significantly predicts the decision to enter the tournament (logit, $p = 0.347$ and $p = 0.593$, respectively). This indicates that the choice between the individual piece rate and the tournament scheme is not systematically driven by ability differences in our sample.

To further examine potential mechanisms underlying tournament entry, we follow Niederle and Vesterlund (2007) and assess whether differences in overconfidence, risk preferences, or feedback aversion may account for the observed gender gap. While feedback aversion cannot be directly tested, relative performance feedback is largely absent in our design: participants receive no information about others' performance or tournament outcomes until the end of the experiment. Consequently, aversion to immediate relative feedback is unlikely to play a central role in the tournament-entry decision.

Table 5: Logistic Regression on Choice of Tournament in Stage 3

	(1)	(2)	(3)	(4)
Dependent variable: Tournament				
Time spent (Stage 1)	−0.006	−0.008	−0.005	−0.007
	(0.007)	(0.007)	(0.007)	(0.007)
Time spent (Stage 2)	−0.004	−0.003	−0.002	−0.002
	(0.007)	(0.007)	(0.007)	(0.007)
Gender (Female = 1)	−0.853*	−0.550	−0.649	−0.435
	(0.502)	(0.531)	(0.523)	(0.543)
Risk attitudes		0.188*		0.158
		(0.104)		(0.106)
Overconfidence			0.360*	0.313
			(0.191)	(0.197)
Constant	1.503	0.153	0.668	−0.339
	(1.179)	(1.416)	(1.242)	(1.431)
Observations	100	100	100	100

Notes: Logit regression with *choice of tournament* as the dependant variable. Columns (2) and (3) add risk attitudes and overconfidence separately. Column (4) includes both controls simultaneously. Standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

We next control for individual risk attitudes and overconfidence in a logistic regression of tournament entry (see Table 5). When included separately, both risk attitudes and overconfidence are positively and marginally significantly associated with choosing the tournament. In particular, overconfidence enters with a positive coefficient that is significant at the 10 percent level, while risk attitudes are also marginally significant when considered on their own.[6] When both controls are included simultaneously, however, neither variable remains statistically significant, suggesting that they capture overlapping variation in tournament entry.

Across all specifications, the coefficient on gender is negative, indicating that women are less likely to enter the tournament than men. However, once risk attitudes and/or overconfidence are included, the gender coefficient decreases in magnitude and becomes statistically insignificant. This pattern suggests that gender differences in tournament entry are at least partly mediated by individual characteristics related to risk-taking and overconfidence, rather than reflecting pure preferences for competition.

Table 6: Payoff Scheme Choice in Stage 3 by Gender

Gender	N	Payoff Scheme Choice	Std. Error	95% CI
Male	50	0.32	0.065	[0.191, 0.449]
Female	50	0.18	0.054	[0.073, 0.286]
Difference (M – F)		0.14	0.085	[–0.027, 0.307]

Notes: Difference tested using a two-sample test of proportions. One-tailed $p(Ha : Male > Female) = 0.053$ Two-tailed $p = 0.106$. Confidence intervals are two-sided.

[6]We use the self-confidence measure explained in the previous section. Results of this section are consistent also if we use only the ranking beliefs (GQ2 answers) instead of the universal self-confidence.

Having established that performance measures do not systematically predict tournament entry and that individual traits such as risk attitudes and overconfidence account for part of the observed gender gap, we proceed to use tournament choice in Stage 3 as an indicator of individual competitiveness. Table 6 Compares entry rates across genders. We find that men are more likely to select the tournament scheme than women: 32% of men choose the tournament, compared to 18% of women, implying a 14-percentage-point gap. This difference is marginally statistically significant (one-tailed proportion test, $p = 0.053$). We state the first result as follows:

> *Result 1:* Men are more likely than women to enter competitive environments, a difference that appears to be partly driven by higher levels of overconfidence and partly by risk preferences.

5.2 Goal Achievement

In Stage 4, 66% of players in the Near treatment meet their personalized goal, compared to 44% in the Far treatment. There is no surprise that the Near threshold is more achievable; however, we examine whether relatively more achievable goals induce larger increases in effort than goals that are set further away and are therefore harder to reach. To measure the change in effort, we look at the difference between the time spent in Stage 4 and the time spent in the baseline Stage 1.

The average time spent in Stage 4 improves relative to Stage 1 for both Near and Far treatments. However, this improvement is substantially larger in the Near treatment (see Table 7). Players in the Near treatment improved by approximately 26 seconds, whereas those in the Far treatment improved by approximately 11 seconds, implying a difference of roughly 15 seconds between the two treatments. This improvement is significantly larger for the Near treatment (t-test, one-tailed p = 0.046), consistent with our pre-registered directional hypothesis. Although the baseline performance in Stage 1 is better for the Near group compared to the Far group, this initial difference is not statistically significant. (See Figure 3)

> *Result 2:* In the presence of the threshold, performance improvement is larger in the treatment with a Near threshold compared to the treatment with Far threshold.

Table 7: Improvement from Stage 1 to Stage 4 by Treatment

Group	N	Mean Improvement	Std. Error	95% CI
Near	50	26.43	6.39	[13.58, 39.27]
Far	50	11.32	6.15	[–1.03, 23.67]
Difference (Near – Far)		15.11	8.87	[–2.49, 32.70]

Notes: Difference tested using a two-sample test of proportions. One-tailed $p(Ha : Near > Far) = 0.0458$ Two-tailed $p = 0.0916$. Confidence intervals are two-sided.

We now look at gender differences in goal achievement and the extent to which these differences may relate to underlying competitiveness. In this part of the analysis, we first document gender gaps in meeting the Stage 4 threshold and in performance improvements

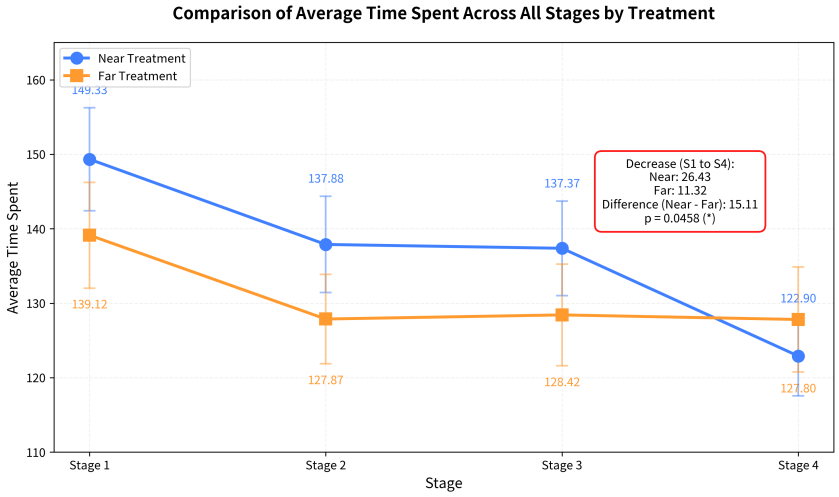

Figure 3: Average time spent across all stages by Treatments. Bars are Standard Errors.

from Stage 1 to Stage 4. The role of competitiveness in explaining these patterns will be addressed afterward and formally explored through a mediation model in the econometric section.

Consistent with our pre-specified directional hypothesis, men meet the Stage 4 goal at a significantly higher rate than women (Table 8). Pooling the Near and Far treatments, 66% of men achieve their personalized threshold compared to 44% of women, a 22-percentage-point difference. This gap is statistically significant under a one-tailed two-sample test of proportions ($z = 2.21$, $p = 0.014$), which reflects the directional hypothesis stated ex ante. This result provides clear evidence of a gender difference in goal attainment within the threshold environment.

> *Result 3:* Men are more likely to achieve their personalized goal in the presence of an incentivized threshold.

Table 8: Goal Achievement by Gender

Group	N	Proportion Achieved	Std. Error	95% CI
Male	50	0.66	0.067	[0.53, 0.79]
Female	50	0.44	0.070	[0.30, 0.58]
Difference (M – F)		0.22	0.097	[0.03, 0.41]

Notes: Difference tested using a two-sample test of proportions. One-tailed $p(Ha : M > F) = 0.0135$ Two-tailed $p = 0.027$. Confidence intervals are two-sided.

The performance data similarly show that men exhibit a larger improvement between Stage 1 and Stage 4. On average, men improve by 28.31 seconds, while women improve by 9.44 seconds, a difference of approximately 18.87 seconds. A one-tailed t-test, appropriate for this directional hypothesis, confirms that this difference is statistically significant ($p = 0.0176$). Thus, not only are men more likely to meet their threshold, but they also display greater gains in performance when the threshold incentive is introduced. (see Figure 4)

Result 4: In the presence of the threshold, men improve their performance more than women

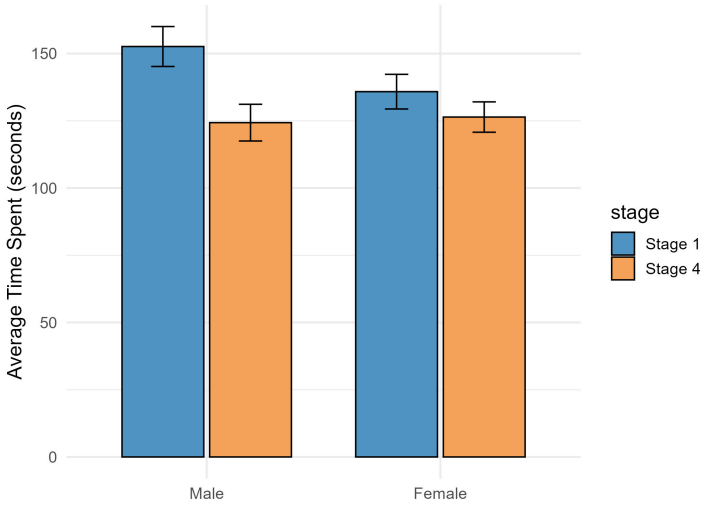

Figure 4: Average time spent in Stage 1 to Stage 4 by gender. Bars are Standard Errors.

These patterns raise the question of whether gender differences in competitiveness contribute to the observed gender gaps in goal achievement and performance improvement. Next, we examine whether individuals' competitiveness, as measured by their choice of payoff scheme in Stage 3, is related to their success in meeting the threshold and whether this mechanism helps explain the gender differences documented here.

Table 9: Goal Achievement by Decision in Stage 3

Decision in Stage 3	N	Proportion Achieved	Std. Error	95% CI
Piece-rate	75	0.507	0.058	[0.39, 0.62]
Tournament	25	0.680	0.093	[0.49, 0.86]
Difference (P – T)		-0.173	0.109	[-0.39, 0.042]

Notes: Players divided into Piece-rate and Tournament groups based on their choice of payoff scheme in Stage 3. Difference tested using a two-sample test of proportions. One-tailed $p(Ha : NC < C) = 0.0657$. Two-tailed $p = 0.131$. Confidence intervals are two-sided.

Goal achievement rates are indeed higher among players who select the tournament in Stage 3: 68% of individuals who choose the tournament achieve their goal compared to 51% of those who choose the individual scheme (Table 9). The difference of 17 percentage points is marginally significant using a one-tailed proportion test ($p = 0.065$), which is appropriate given the directional nature of the hypothesis. While the effect does not reach conventional 5% significance levels, the result is aligned with the prediction that competitiveness is positively associated with goal attainment. Therefore, we state:

Result 5: Choice of tournament in Stage 3 is marginally and positively correlated with goal achievement

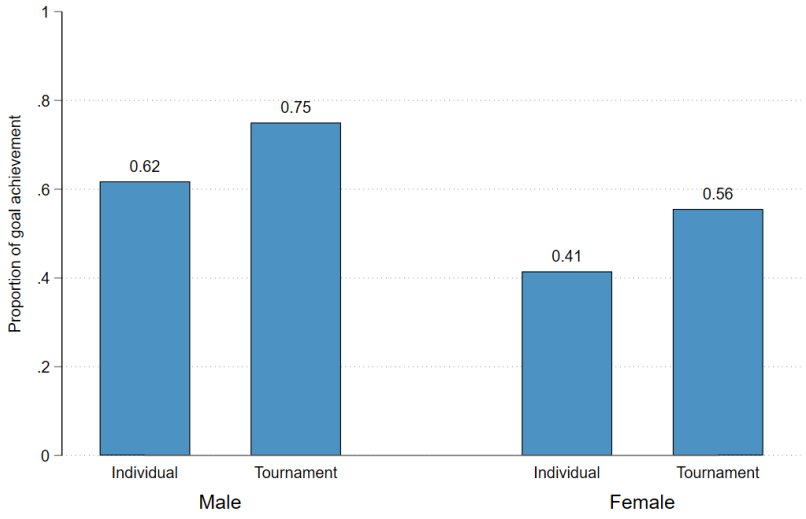

Figure 5: Proportion of goal achievement disentangled by decision in Stage 3 and by gender. *Individual* is Piece-rate.

The relationship between goal achievement and tournament entry is disentangled for men and women in Figure 5. Players who choose the tournament have higher rates of goal achievement, and men are performing better than women in achieving personalized goals; however, we can see from the figure that men perform better than women also among the players who chose the piece-rate (individual) payoff scheme. This relationship will be examined more rigorously in the following section through a structural equation model that accounts for gender, treatment assignment, and baseline performance.

5.3 Econometric Model

To examine whether the effect of gender on goal achievement operates partly through differences in competitiveness, a mediation model was estimated using structural equation modeling (SEM). The model included gender as the exogenous predictor, competitiveness as the mediating variable, and goal achievement as the outcome variable. Since both the mediator and the outcome are binary, the SEM is specified as a linear probability mediation model, which allows for a transparent decomposition of direct and indirect effects.[7]

The system of equations for the model can be expressed as:

$$\text{Competitiveness}_i = \alpha_1 + a\,\text{Gender}_i + \gamma_j\,\text{Performance}_{j,i} + e_{1i} \tag{16}$$

$$\text{GoalAchieved}_i = \alpha_2 + b\,\text{Competitiveness}_i + c'\,\text{Gender}_i + \delta\,\text{Treatment}_i + e_{2i} \tag{17}$$

Where *Competitiveness* is a binary indicator equal to 1 if participant i chooses the tournament scheme in Stage 3 and 0 if she chooses the piece-rate scheme. *Gender* is a binary variable equal to 1 for female participants. $Performance_{j,i}$ denotes performance

[7]Estimating the model using a generalized structural equation model with logit links for the binary variables yields nearly identical coefficient estimates, confirming that the results are not driven by the linear probability specification.

measures from earlier stages, indexed by $j \in \{1, 2\}$, where performance is measured by the time spent to complete the task in Stage 1 and Stage 2, respectively. *GoalAchieved* is a binary indicator equal to 1 if participant i meets the personalized performance threshold in Stage 4 and 0 otherwise. *Treatment* is a binary variable indicating assignment to the Far threshold condition. Finally, e_{1i} and e_{2i} denote error terms in the mediator and outcome equations, respectively.

The equation for the mediator (16) specified competitiveness as a function of gender and two control variables, performance in Stage 1 and performance in Stage 2 (measured with the time spent). The outcome equation (17) modeled goal achievement as a function of competitiveness, gender, and treatment assignment. All variables were observed; no latent constructs were specified. Following standard practice, estimation relies on maximum likelihood with normally distributed latent disturbances. Results are presented in Table 10 and summarized below.

Table 10: Structural Equation Model Estimates for Goal Achievement Across Multiple Specifications

	(1)	(2)	(3)	(4)
Mediator equation: Competitiveness				
Gender (Female = 1)	-0.153^{*}	-0.112	-0.075	-0.112
	(0.088)	(0.088)	(0.092)	(0.088)
Performance Stage 1	-0.01	-0.00	-0.00	-0.00
	(0.001)	(0.001)	(0.001)	(0.001)
Performance Stage 2	-0.00	-0.00	-0.00	-0.00
	(0.001)	(0.001)	(0.001)	(0.001)
Overconfidence		0.064^{**}	0.054^{*}	0.064^{**}
		(0.031)	(0.032)	(0.031)
Risk Attitudes			0.024	
			(0.017)	
Outcome equation: Goal Achievement				
Competitiveness	0.183^{*}	0.183^{*}	0.183^{*}	0.217^{*}
	(0.110)	(0.110)	(0.110)	(0.112)
Gender (Female = 1)	-0.205^{**}	-0.205^{**}	-0.205^{**}	-0.231^{**}
	(0.094)	(0.094)	(0.094)	(0.095)
Overconfidence				-0.052
				(0.035)
Treatment (Far = 1)	-0.254^{***}	-0.254^{***}	-0.254^{***}	-0.251^{***}
	(0.094)	(0.094)	(0.094)	(0.093)
Indirect (Gender → Compet. → Goal)	-0.028	-0.020	-0.014	-0.024
	(0.023)	(0.020)	(0.019)	(0.023)
Total Effect (Gender → Goal)	-0.232^{**}	-0.225^{**}	-0.218^{**}	-0.255^{**}
	(0.094)	(0.094)	(0.095)	(0.096)
Observations	100	100	100	100

Notes: Coefficients from linear probability SEM estimated by maximum likelihood. Robustness specifications add overconfidence and risk attitudes to the mediator equation in columns (2) and (3) and add overconfidence to both the mediator and outcome equations in column (4). Standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

The direct effect (c') captures the effect of gender on goal achievement independent of competitiveness. The indirect effect of gender on goal achievement operates through

the product $a \times b$, representing the pathway Gender $\rightarrow$ Competition $\rightarrow$ GoalAchieved, and the total effect is the sum of the direct and indirect components.

Table 10 reports estimates from the baseline structural equation model and several alternative specifications. Column (1) presents the baseline model including gender and performance controls in the mediator equation. Consistent with earlier results, gender is negatively and marginally significantly associated with competitiveness ($a = -0.153$, $p = 0.083$), indicating that women are less likely than men to choose the tournament payoff scheme, conditional on performance in earlier stages. Performance in Stages 1 and 2 does not significantly predict tournament entry, reinforcing the interpretation of tournament choice as a behavioral measure of competitiveness rather than ability.

Competitiveness is positively associated with goal achievement across all specifications. In the baseline model, competitive individuals are more likely to achieve the performance threshold ($b = 0.183$, $p < 0.098$). This relationship remains stable when additional controls are introduced in Columns (2) through (4), suggesting that the positive association between competitiveness and success is robust.

The direct path from gender to goal achievement is negative and statistically significant in the baseline model ($c' = -0.20, p = 0.030$), suggesting that women are less likely than men to achieve the goal even after accounting for competitiveness and treatment status. Treatment assignment itself had a significant negative effect on goal achievement ($\delta = -0.25, p = 0.007$), meaning that for the Far treatment it is %25 less likely to achieve the goal.

Decomposition of total effects indicated that the indirect effect of gender through competition was negative but not statistically significant ($\beta = -0.03, p = 0.23$). The total effect of gender on goal achievement remained significant ($\beta = -0.23, p = 0.014$). This pattern indicates that competitiveness does not mediate the observed gender gap in goal achievement; instead, most of the gender difference operates through a direct channel. The structure of the mediation model and estimated path coefficients are illustrated graphically in Figure 6.

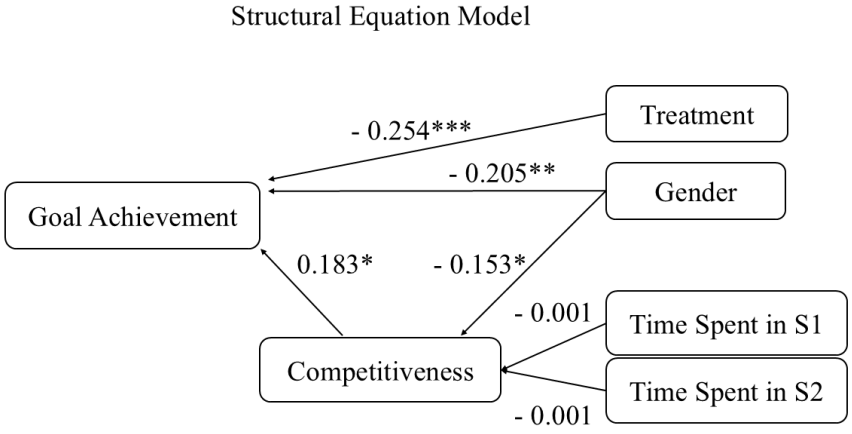

Figure 6: Structural path diagram for the mediation model. Gender predicts both competitiveness and goal achievement, with competitiveness serving as a potential mediator. Coefficients shown on arrows are unstandardized SEM estimates with significance levels indicated by stars ($^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$).

5.4 Robustness Check: Risk Attitudes and Overconfidence

To assess whether the mediation results are sensitive to additional behavioral controls, we re-estimate the structural model including measures of risk attitudes and overconfidence. Columns (2) and (3) of Table 10 add overconfidence and risk preferences to the mediator equation. Overconfidence is positively associated with tournament entry and statistically significant or marginally significant across specifications, whereas risk attitudes are not. Once overconfidence is included, the coefficient on gender in the mediator equation becomes smaller and statistically insignificant, suggesting that part of the gender difference in tournament entry operates through confidence-related traits.

Importantly, the estimated effect of competitiveness on goal achievement remains positive and robust across all specifications. When overconfidence is additionally included in the outcome equation (Column 4), the coefficient on competitiveness increases in magnitude and becomes statistically stronger, indicating that competitiveness predicts goal achievement independently of confidence. Across all models, the direct effect of gender on goal achievement remains negative and statistically significant, confirming that competitiveness and overconfidence do not fully account for the observed gender gap. Overall, these results support the main conclusions and suggest that overconfidence is an important upstream determinant of competitive behavior rather than a confounding factor.

6 Conclusion

This chapter investigates how individuals respond to personalized performance thresholds and how these responses vary by goal proximity, competitiveness, and gender. Using a controlled laboratory experiment combined with a simple theoretical framework, the analysis sheds light on why some individuals increase effort and succeed in reaching performance targets while others do not, even when facing similar incentives.

The experimental results provide clear evidence that thresholds are effective in shaping behavior. Introducing a performance goal leads to significant improvements relative to baseline performance, indicating that the presence of a target can motivate additional effort. However, these effects are not uniform. Participants assigned to more attainable thresholds exhibit both higher rates of goal achievement and larger average performance improvements than those facing more demanding goals. This finding aligns closely with the theoretical model, which predicts that effort adjustments become less attractive as the distance to the threshold increases. Together, the theory and experimental evidence highlight the importance of goal proximity as a key determinant of effort provision.

A central contribution of this chapter is the documentation of substantial gender differences in responses to performance incentives. Men are more likely to select competitive compensation schemes and are more likely to achieve their performance goals. They also exhibit larger performance improvements in the presence of thresholds. These differences persist even though women perform at least as well as men in the baseline task. This pattern suggests that gender gaps in outcomes are not driven by initial ability but emerge through differential responses to incentive structures.

The analysis further explores whether competitiveness can explain these gender differences. Consistent with previous literature, tournament entry serves as a meaningful behavioral measure of competitiveness, as it is not systematically predicted by performance. Competitiveness is positively associated with goal achievement, indicating that

individuals who self-select into competitive environments tend to perform better under threshold incentives. However, mediation analysis using structural equation modeling shows that competitiveness does not account for the observed gender gap in goal achievement. While women are less likely to enter tournaments, the indirect pathway from gender to goal achievement through competitiveness is small and statistically insignificant. Most of the gender difference appears to operate through a direct channel.

Robustness checks examine the roles of risk attitudes and self-confidence. Overconfidence, measured through incentivized belief elicitation, emerges as an important predictor of competitive choices and partially attenuates the gender effect in tournament entry. However, even after accounting for these traits, gender differences in goal achievement remain, indicating that additional psychological or contextual mechanisms are likely at play.

Taken together, the findings underscore that personalized performance goals are not neutral policy tools. While thresholds can motivate effort on average, they also amplify heterogeneity across individuals. In particular, the interaction between goal proximity and individual traits such as gender, competitiveness, and confidence shapes both effort provision and success. These results have important implications for the design of incentive schemes in settings such as education, workplaces, and performance evaluations, where goal-based incentives are widely used. Designing goals that are too demanding may discourage effort among some groups, while more attainable targets can foster engagement and improvement more broadly.

This chapter also opens several avenues for future research. One promising direction is to examine how feedback, learning, and belief updating interact with threshold incentives over time. Another is to explore alternative goal structures, such as flexible or self-chosen thresholds, and their implications for inequality in outcomes. Finally, extending the analysis to field settings would help assess the external validity of the experimental findings and inform the design of equitable and effective incentive systems.

In sum, this chapter demonstrates that the effectiveness of performance thresholds depends not only on their calibration but also on who faces them. Understanding this interaction is crucial for designing incentive schemes that motivate effort without exacerbating existing disparities.

References

Abeler, J., Falk, A., Goette, L., & Huffman, D. (2011). Reference points and effort provision. *American Economic Review*, *101*(2), 470–492.

Albert, D. A., & Smilek, D. (2023). Comparing attentional disengagement between prolific and mturk samples. *Scientific Reports*, *13*(1), 20574.

Apicella, C. L., Demiral, E. E., & Mollerstrom, J. (2020). Compete with others? no, thanks. with myself? yes, please! *Economics Letters*, *187*, 108878.

Charness, G., Gneezy, U., & Henderson, A. (2018). Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior & Organization*, *149*, 74–87.

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

Chiapello, M. (2018). Balancetable: Stata module to build a balance table and print it in a latex file or an excel file.

Corgnet, B., Gómez-Miñambres, J., & Hernán-Gonzalez, R. (2015). Goal setting and monetary incentives: When large stakes are not enough. *Management Science*, *61*(12), 2926–2944.

De Paola, M., Gioia, F., & Scoppa, V. (2015). Are females scared of competing with males? results from a field experiment. *Economics of Education Review*, *48*, 117–128.

Falk, A., Becker, A., Dohmen, T., Huffman, D., & Sunde, U. (2023). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Management Science*, *69*(4), 1935–1950.

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

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The quarterly journal of economics*, *118*(3), 1049–1074.

Horn, D., Kiss, H. J., & Lénárd, T. (2022). Gender differences in preferences of adolescents: Evidence from a large-scale classroom experiment. *Journal of Economic Behavior & Organization*, *194*, 478–522.

Lénárd, T., Horn, D., & Kiss, H. J. (2024). Competition, confidence and gender: Shifting the focus from the overconfident to the realistic. *Journal of Economic Psychology*, *104*, 102746.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American psychologist*, *57*(9), 705.

Markowsky, E., & Beblo, M. (2022). When do we observe a gender gap in competition entry? a meta-analysis of the experimental literature. *Journal of Economic Behavior & Organization*, *198*, 139–163.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The quarterly journal of economics*, *122*(3), 1067–1101.

Palan, S., & Schitter, C. (2018). Prolific. ac—a subject pool for online experiments. *Journal of behavioral and experimental finance*, *17*, 22–27.

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior research methods*, *54*(4), 1643–1662.

Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the big five inventory–2: The bfi-2-s and bfi-2-xs. *Journal of Research in Personality*, *68*, 69–81.

Van Veldhuizen, R. (2022). Gender differences in tournament choices: Risk preferences, overconfidence, or competitiveness? *Journal of the European Economic Association*, *20*(4), 1595–1618.

7 Appendix

Experiment Instructions

Below is the content of the experiment as displayed on subjects' screens.

SCREEN 1

Displays the Participants Information Sheet and Consent Form according to the format provided by the Ethics Committee. Participants begin the study only if they agree to the consent form, and exit the study otherwise.

SCREEN 2

WELCOME! Thank you for taking part in this study!
Please, read all instructions carefully to make sure you understand the rules of the study and how your payment will be determined.
You will not be allowed to go back to previous pages. Your earnings will depend on your decision, other participants' decisions, and chance.
You will not know the identities of any participants and they will not know yours.
Individual payments will remain private, as nobody will know the other participants' payments.
You can abandon the study at any time. However, if you abandon the study before its completion, you will not receive any payment.
You will receive the payment according to Prolific rules, within a reasonable time frame from the completion of the study.

When you are ready to start today's study, please click [NEXT] to continue.

SCREEN 3

INSTRUCTIONS

In the experiment today you will be asked to complete four different tasks, after some tasks you will be asked additional "Guessing questions". After completing all tasks, we will ask you to answer a very short survey.

Each task will take about 5 minutes to complete.

You will get a participation fee of 5.00 pounds. In addition, we will randomly select one of the tasks and pay you based on your performance in that task. We also randomly select one of the guessing questions and pay you for a correct answer to that question. The random selection of the guessing question is not related to the random selection of a task for payment.

The method we use to determine your earnings varies across tasks. Before each task, we will describe in detail how your payment is determined. Please pay attention to each instruction set.

Therefore, your final payment will include:

- A participation fee of 5.00 pounds;
- The earnings from the randomly selected task;
- The earnings from the randomly selected guessing question.

During the study, your earnings will be denoted in points. The exchange rate is: 20 points = 1.00 pounds.

In the next screens, and throughout the study, you will be asked some comprehension questions about the instructions you read. At the end, there will also be a short questionnaire. These questions will not provide you with additional earnings, but their completion is necessary to proceed to the task.

Your answers to all tasks and questions are anonymous.

SCREEN 4

Comprehension Questions:

CQ1. You will be paid for all the tasks you perform during the study.

TRUE / FALSE → FALSE

In case of wrong answer provided, the following text is shown:

Remember: Once you have completed the four tasks, the computer will randomly select one of the four tasks and we pay you based on your performance in that task.

CQ2. You will be paid for all the guessing questions you answer correctly during the study.

TRUE / FALSE → FALSE

In case of wrong answer provided, the following text is shown:

Remember: Once you have completed the four tasks, the computer will randomly select one of the three guessing questions and we pay you if you answered that question correctly.

CQ3. Your total earnings are based only on the participation fee and your performance on the tasks.

TRUE / FALSE → FALSE

In case of wrong answer provided, the following text is shown:

Remember: you have the chance to earn additional earnings by answering the guessing questions correctly. Only one of the three guessing questions will randomly be chosen for payment.

SCREEN 5 – Stage 1

Task 1 Instructions

You have 5 minutes to solve 5 tasks. The faster you correctly complete all tasks, the higher your payoff.

Your task is to count how many times a specific letter appears in a grid box filled with random letters.

To submit your answer:

- Type the number into the box below the grid.
- Click the Submit button.

After each submission:

- A new letter grid will appear.
- A new target letter will be provided.

At the top of the screen, you can track:

- Your number of correct and incorrect answers.
- The time remaining.

Your Payoff
If Task 1 is the one randomly selected for payment, your payoff is:

- If you correctly complete the task within 5 minutes, your payoff is your time left (seconds left out of 5 minutes) + 5 points.
  Example: 30 seconds are left from the timer after you have submitted 5 correct answers, then your payoff is 30 + 5 = 35 points.
- If the 5-minute time limit ends and you have not submitted 5 correct answers, you will receive 1 point per correct answer you have submitted.
  Example: You submit 3 correct answers and the time limit ends, then your payoff is 3 points.
- If you hit the limit of wrong submissions, the task stops and your payoff is zero. The limit of wrong submissions is 5.
  If you submit 5 wrong answers, you earn no payoff regardless of your correct submissions or your time left.

We refer to this payment as the "individual" payment.

In any of the cases above, you can still continue the study for the following stages. However:
Exclusion criteria: Participants who do not submit any correct answers within the 5-minute period will be excluded from the study and will not receive the participation fee.

You will practice the described task. In this practice round, you have 3 minutes to complete 3 letter grids. The same rules as the main round apply, but this round does not affect your final payment. The "Exclusion criteria" does not apply to the practice round.

In the next screens, you will be asked some comprehension questions about the instructions you just read.

SCREEN 6 – Stage 1

Comprehension Questions:

CQ4. In the main round, you have 5 minutes to submit five correct answers.

TRUE / FALSE → TRUE

In case of wrong answer provided, the following text is shown:

Remember: In the main round, you have 5 minutes to submit 5 correct answers. However, the goal is to do this as fast as you can.

CQ5. If you submit 5 wrong answers, the game ends and you will only earn points from your correct answers.

TRUE / FALSE → FALSE

In case of wrong answer provided, the following text is shown:

Remember: If you submit 5 wrong answers, the game will end, and your additional payoff will be zero (you will receive only the participation fee).

CQ6. The time you take to complete the task, does not impact your payoff.

TRUE / FALSE → FALSE

In case of wrong answer provided, the following text is shown:
Remember: The faster you correctly complete all tasks, the higher your payoff.
CQ7. The game involves always counting the same letter in a series of images.
TRUE / FALSE → FALSE
In case of wrong answer provided, the following text is shown:
Remember: The game involves counting how many times a specific letter appears in a series of images. After each submission, a new set of letters will appear and a new target letter will be provided.
GET READY, On the next page, the practice round will begin.
Your 3-minute countdown will start when you click to go to the next page.

SCREEN 7 – Stage 1

Practice round

SCREEN 8 – Stage 1

Practice Feedback
You solved 3 tasks correctly! // You reached the maximum limit of incorrect answers! // Your time is up!

- Correct answers: {{ correct_count }}
- Wrong answers: {{ wrong_count }}
- Time spent: {{ time_spent }} seconds

This was a practice round. Your performance doesn't affect your final payoff.
GET READY! The main round begins on the next page.
Your 5-minute countdown will start when you click to proceed.
Remember: Your performance in this round will affect your final payoff.

SCREEN 9 – Stage 1

Grid task

SCREEN 10 – Stage 1

Feedback
You solved 5 tasks correctly! // You reached the maximum limit of incorrect answers! // Your time is up!

- Correct answers: {{ correct_count }}
- Wrong answers: {{ wrong_count }}
- Time spent: {{ time_spent }} seconds

SCREEN 11 – Stage 1

Guessing question 1:

Do you think you performed above/below/similar to the average participant?
If this question is randomly chosen for payment at the end of the experiment, then
- if you guessed correctly, you will earn an additional 10 points.
- If you did not guess correctly no additional amount will be added or deducted to your total points.

- Above
- Similar to (i.e., same number of correct answers within 10 seconds above or below the average)
- Below

SCREEN 12 – Stage 2

Task 2 Instructions

You have 5 minutes to submit 5 correct answers in the task, which is the same as Task 1. The faster you correctly complete the task, the higher your potential payoff.

Your task is to count how many times a specific letter appears in a grid box filled with random letters, but unlike Task 1, your payment depends on your performance compared to a group of four participants (including you), randomly selected from study participants. You won't know their identities, and they won't know yours.

Your Payoff

If Task 2 is randomly selected for payment, your payoff depends on your ranking within your group of four participants. Rankings are determined as follows:

- Players who submit 5 correct answers are ranked by how fast they finish the task (from fastest to slowest).
- Players who do not submit 5 correct answers are ranked below those who finish, based on their number of correct answers (from highest to lowest).

Example: Players A and B submit 5 correct answers in 250 and 280 seconds, respectively. Player C submits 4 correct answers, and Player D submits 3. Players A and B rank 1st and 2nd (A is faster), Player C ranks 3rd, and Player D ranks 4th.

- The 1st-ranked participant receives a payoff of 4 × (seconds left out of 5 minutes + 5 points).
  Example: If you rank 1st with 30 seconds left after submitting 5 correct answers, your payoff is $4 \times (30 + 5) = 140$ points.
- Other group members (ranks 2–4) receive 0 points.
- If there's a tie for 1st, the winner is randomly chosen among those tied.

We refer to this as the "tournament" payment.

You will not be informed of your ranking in the tournament until all four tasks are completed.

After performing the task, you will self-assess your performance.

In the next screens, you will be asked some comprehension questions about the instructions you just read.

SCREEN 12 – Stage 2

Comprehension Questions:

CQ8. You will be paid only if you rank first in your four people group.

TRUE / FALSE → TRUE

In case of wrong answer provided, the following text is shown:

Remember: The individual that ranks first in the group, will receive four times the payoff. The other group members receive no payment.

CQ9. The player who submits 5 correct answers in the lowest time in the group, will receive 40 points in total for this stage.

TRUE / FALSE → FALSE

In case of wrong answer provided, the following text is shown:

Remember: The individual that ranks first in the group, will receive four times the payoff, that is 4 * (time left + 1 point per correct image). The other group members receive no payment (0 points).

GET READY! The main round begins on the next page.

Your 5-minute countdown will start when you click to proceed.

SCREEN 13 – Stage 2

Grid task

SCREEN 14 – Stage 2

You solved 5 tasks correctly! // You reached the maximum limit of incorrect answers! // Your time is up!

Your ranking in the group will be revealed to you at the end of the study.

SCREEN 15 – Stage 2

Guessing question 2: What is your best guess about your ranking in your group for this task? (Where 1 is the person with the highest number of images solved and lowest time taken and 4 is the one with the lowest number of images and/or highest time taken.)

If this question is randomly chosen for payment at the end of the experiment, then
- if you guessed correctly, you will earn an additional 10 points.
- If you did not guess correctly no additional amount will be added or deducted to your total points.

- 1
- 2
- 3
- 4

SCREEN 16 – Stage 3

Task 3 Instructions

As in the previous tasks, you have 5 minutes to submit 5 correct answers in the task. The faster you correctly complete the task, the higher your potential payoff.

Your task is to count how many times a specific letter appears in a grid box filled with random letters. However, in this task, you can choose between the "individual" or "tournament" payment scheme for this task.

Your Payoff

If Task 3 is randomly selected for payment, your payoff depends on your chosen payment scheme:

Individual Payment

- If you submit 5 correct answers within 5 minutes, your payoff is seconds left out of 5 minutes + 5 points.
  Example: If you submit 5 correct answers with 30 seconds left, your payoff is 30 + 5 = 35 points.
- If you submit fewer than 5 correct answers within 5 minutes, your payoff is 1 point per correct answer.
  Example: If you submit 3 correct answers, your payoff is 3 points.
- If you submit 5 incorrect answers, the task stops, and your payoff is 0 points.

Tournament Payment

Your performance is compared to the same group of four participants (including you) from Task 2. Rankings are determined as follows:

- Players who submit 5 correct answers are ranked by how fast they finish the task (from fastest to slowest).
- Players who do not submit 5 correct answers are ranked below those who finish, based on their number of correct answers (from highest to lowest).
- The 1st-ranked participant receives a payoff of 4 × (seconds left out of 5 minutes + 5 points), even if tied for 1st.
  Example: If you rank 1st with 30 seconds left after submitting 5 correct answers, your payoff is 4 × (30 + 5) = 140 points.
- Other group members (ranks 2–4) receive 0 points.
- If there's a tie for 1st, all tied players receive the 1st-place payoff.

In the next screens, you will be asked some comprehension questions about the instructions you just read.

SCREEN 17 – Stage 3

CQ10. If you solve 5 images in the lowest time than the other group members did in Task 2, you will receive 40 points in total for this stage, regardless of whether you choose the "piece rate" or "tournament" option.

TRUE / FALSE → FALSE

In case of wrong answer provided, the following text is shown:

Remember: If you rank first (based on number of correct images, and time taken) compared to the other group members did in Task 2, you will receive four times the payoff, only if you select the "tournament" option. If you select the "individual rate" option, you receive points equal to the number of correct answers you submit + your remaining seconds, regardless of what others have done.

SCREEN 18 – Stage 3

Task 3: Your payoff choice Select your preferred payment option:

- Individual
- Tournament

Reminder:
Individual payment: Time left (in seconds) out of 5 minutes plus 1 point for each correct answer.
Tournament payment: Four times your individual payment if you rank first in your group; otherwise, 0 points.

GET READY! The main round begins on the next page.

Your 5-minute countdown will start when you click to proceed.

SCREEN 19 – Stage 3

Grid task

SCREEN 20 – Stage 3

Task ended!

You will be informed of Task 3's results only at the end of the session.

SCREEN 21 – Stage 4

Note to reader – X refers to the individual goal of participants: if they did not manage to get 5 correct images in Stage 1, X is the highest number of correct images they obtained. If they had 0 correct images in Stage 1, they had been removed from the experiment immediately after stage 1.
Note to reader - Time threshold αt is based on (t), player's time spent in Stage 1 to submit their last correct answer, and a multiplier $\alpha = 0.95$ for the Near treatment and $\alpha = 0.90$ for the Far treatment.
Task 4 Instructions
As in the previous tasks, you have 5 minutes to submit 5 correct answers in the task. The faster you correctly complete the task, the higher your potential payoff. In this task you can earn an extra 20 points if you can meet the threshold given below.

Your task is to count how many times a specific letter appears in a grid box filled with random letters.
Threshold
Your threshold goal is to submit X correct answers in αt seconds or less.

In the first stage, you correctly solved X letter grids in t seconds.

Your Payoff
If Task 4 is randomly selected for payment, your payoff is calculated with the "individual payoff" scheme as in Task 1, with an additional 20 points if you meet the threshold of X correct answers within αt seconds:

1. If you correctly complete the task within 5 minutes, your payoff is seconds left out of 5 minutes + 5 points.
   If you also meet your threshold (X correct answers within αt seconds), your payoff is seconds left out of 5 minutes + 5 points + 20 points.

2. If the 5-minute time limit ends and you have not submitted 5 correct answers, you will receive 1 point per correct answer submitted.
   If you still meet your threshold (X correct answers within αt seconds), your payoff is 1 point per correct answer submitted + 20 points.

3. If you submit 5 wrong answers, you have hit the limit, the task stops, and your payoff is 0 points.
   If you meet your threshold (X correct answers within αt seconds) before hitting the limit, your payoff is 20 points.

Note: Payoff cases 2 and 3 are valid only for players whose threshold is less than 5 correct answers.

Before performing the task, you will be asked to predict your performance.

In the next screens, you will be asked some comprehension questions about the instructions you just read.

SCREEN 22 – Stage 4

CQ11. You will receive an extra 20 points if you achieve the goal given to you within the time threshold asked.

TRUE / FALSE → TRUE

In case of wrong answer provided, the following text is shown:

Remember: if your threshold is to solve X images in the given time, and you correctly solve X images within that time, you will receive 20+ seconds remaining + X points. But if you correctly solve X images or less, but in more time than the given time, you will receive seconds remaining + number of correct images points.

SCREEN 23 – Stage 4

Guessing question 3: Do you think you will meet the threshold? (Yes/No)

If this question is randomly chosen for payment at the end of the experiment, then
-if you guessed correctly, you will earn an additional 10 points.
-If you did not guess correctly, no additional amount will be added or deducted to your total points.

GET READY! The main round begins on the next page.

Your 5-minute countdown will start when you click to proceed.

SCREEN 24 – Stage 4

Grid task

SCREEN 25 – Post-experiment questionnaire

Thank you for completing the study!

We will now ask you some additional and demographic questions. Your answers from this point onwards will not affect your payment in any way. However, please complete the survey so that you can proceed to payment.

SCREEN 26 – Post-experiment questionnaire

Hereby you can find a series of characteristics that may or not describe you. For example, do you agree that you are a person that enjoys spending time with others?

Indicate next to each affirmation how much you agree or disagree:

I see myself as someone who... (Strongly agree [1] – Strongly disagree [7])

- ... is reserved
- ... is generally trusting
- ... tends to be lazy
- ... is relaxed, handles stress well
- ... has few artistic interests
- ... is outgoing, sociable
- ... tends to find fault with others
- ... does a thorough job
- ... gets nervous easily
- ... has an active imagination

Choose the scale to which the following statements describe you: (7-point Likert scale from 1 – "Not at all like me" to 7 – "Exactly like me")

- Competition brings the best out of me, especially if I'm competing against others
- Competition brings the best out of me, especially if I'm competing against myself, without other competitors involved

In general, how willing or unwilling are you to take risks? (scale from 0 [completely unwilling to take risks] to 10 [very willing to take risks])

1. Age

What's your age?

2. Chosen gender / gender identity

What is your current gender?

(Male / Female / Non-binary / Transfeminine / Transmasculine / Genderfluid / Other / Rather not say)

3. What is the highest grade or year of school you completed?

- Never attended school or only kindergarten
- Year 1 through 6 (Primary School)
- Year 7 through 9 (Some Secondary School)
- Year 10 through 11 or GCSE (High School Graduate)
- Year 12 through 13 (A Levels)
- University/College 1 year to 3 (Undergraduate Study or Vocational School)
- University 4 years or more (Postgraduate Study)
- PhD
- Refuse to answer

4. Are you currently. . . ?

- Employed for wages
- Self-employed
- Out of work for 1 year or more
- Out of work for $<$ 1 year
- A homemaker
- A student
- Retired
- Unable to work
- Refuse to answer

5. What is your field of study or occupation?

- Economics
- Other Social Sciences (e.g., Psychology, Sociology, Political science)
- Humanities (e.g., Philosophy, Literature, Linguistics and languages)
- Natural Sciences (e.g., Biology, Chemistry, Physics)
- Formal Sciences (e.g., Computer Sciences, Systems Science, Mathematics)
- Law
- Other professions and applied sciences (e.g., Agriculture, Business, Medicine)
- Other

6. Experience in similar experiments
Have you taken part previously in similar experiment?
Yes / No
You are almost finished with the HIT.

Note that there is no penalty for your response to any question, so please answer honestly. Your responses cannot be used against you in any way.

7. Have you ever misrepresented yourself in order to complete a study on Prolific (such as, on your profile)?
Yes
No

8. Have you ever completed a study on Prolific with a lack of effort, attention, with nonsensical responses, or by "clicking through"?
Yes
No

9. Please indicate how much you agree with the following statements:
I made each decision in this study carefully
Strongly disagree · · · Strongly agree *(Scale 1 to 5)*
I made decisions in this study randomly
Strongly disagree · · · Strongly agree *(Scale 1 to 5)*

SCREEN 27 – Payment

The study has ended. Thank you for your participation.

Here below, you will find a summary of your performance and answers. The payment will be computed after the session has ended for all participants. You will receive the additional payment as bonus through Prolific.

Below are the performances in each Task:

Task 1:

- *If submitted correct answers:*
  You submitted [number of correct answers] correct and [number of wrong answers] wrong answers in [time spent] seconds.
- *If hit the maximum wrong answers:*
  You hit the maximum wrong answers!
- *If ran out of time:*
  You submitted [number of correct answers] correct and [number of wrong answers] wrong answers and ran out of time!

Task 2:
You submitted [number of correct answers] correct and [number of wrong answers] wrong answers in [time spent] seconds.

- *If ranking is pending:*
  Your ranking is not available yet because some group members have not completed Task 2.
- *Otherwise*
  Your rank in Task 2 was: [Task 2 rank]
  Winner: Yes/No

Task 3:
You submitted [number of correct answers] correct and [number of wrong answers] wrong answers in [time spent] seconds.
Your payoff choice was: [individual / tournament]

- *If individual: payoff not shown here.*
- *If tournament*
  Your rank in Task 3 was: [Task 3 rank]
  Winner: Yes/No

Task 4:
You submitted [number of correct answers] correct and [number of wrong answers] wrong answers in [time spent] seconds.

- *If goal achieved:*
  You did meet the threshold set.
- *Otherwise*
  You did not meet the threshold set.

Guessing Question 1: *Player's answer displayed.*
Guessing Question 2: *Player's answer displayed.*
Guessing Question 3: *Player's answer displayed.*

Bonus Selection
The computer has randomly drawn one of the four tasks to determine your earnings. The task randomly drawn is [randomly selected task].

The computer has also randomly drawn one of the three guessing questions to determine your earnings. The guessing question randomly drawn is [randomly selected guessing question].

Within one week, following Prolific rules, you will be paid a bonus on top of the participation fee of 6 pounds.

Please proceed to the next page to complete the study.

SCREEN 28 – Closing page

Thank you for having completed the study!

You will receive your payment within the next few days.

You will be automatically redirected to Prolific in a few seconds.