

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Worker autonomy and performance: Evidence from a real-effort experiment

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:

Rattini, V. (2023). Worker autonomy and performance: Evidence from a real-effort experiment. JOURNAL OF ECONOMICS & MANAGEMENT STRATEGY, 32(2), 300-327 [10.1111/jems.12511].

Availability:

This version is available at: <https://hdl.handle.net/11585/920485> since: 2023-09-04

Published:

DOI: <http://doi.org/10.1111/jems.12511>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).
When citing, please refer to the published version.

(Article begins on next page)

This is the final peer-reviewed accepted manuscript of:

Rattini, V. (2022). Worker autonomy and performance: Evidence from a real-effort experiment. *Journal of Economics & Management Strategy*.

The final published version is available online at:

<https://doi.org/10.1111/jems.12511>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>)

When citing, please refer to the published version.

Worker Autonomy and Performance: Evidence from a Real-Effort Experiment

Veronica Rattini [†]

Abstract

Worker flexibility in effort allocation is a crucial factor for productivity and optimal job design. This paper runs a real-effort experiment that manipulates both the degree and type of autonomy individuals have in scheduling their effort, and it examines the causal effects of these manipulations on final performance. The main findings come from comparing subjects with different levels of cognitive ability. Using individual data on scheduling decisions, I find significant baseline differences in performance and effort allocation strategies between high- and low-cognitive ability subjects. Moreover, the experiment shows that high-ability individuals reach higher performance when they have full scheduling flexibility while limiting any task-ordering possibility increases the performance of low-ability individuals. Overall, this paper provides new and robust evidence on the importance of cognitive ability in explaining effort allocation decisions, and it identifies job design interventions to increase the performance of high- and low-ability workers.

Keywords: Autonomy, Performance, Multitasking, Deadlines, Cognitive Ability.

JEL classification: J20, C91, C80.

[†]Department of Economics, University of Milano, Milano, 20122, Italy. Contact author: veronica.rattini@unimi.it I would like to thank Maria Bigoni, Marco Casari, Guillaume Fréchet, Margherita Fort, Osea Giuntella, David Huffman, Andrea Ichino, George Loewenstein, Ernesto Reuben, Andrew Schotter, Roberto Weber, who gave useful advice in several discussions. I am also grateful to the seminar participants at the University of Pittsburgh, University of Bologna and New York University, and to the attendants of the Workshop on Choice-Process Data in Experimental Economics at the 2018 North-American ESA Conference. This is the accepted version of the following article: Rattini, V. (2023). Worker autonomy and performance: Evidence from a real-effort experiment. *Journal of Economics & Management Strategy*, 1 28., which has been published in final form at <https://doi.org/10.1111/jems.12511>. This article may be used for noncommercial purposes in accordance with the Wiley Self-Archiving Policy.

1 Introduction

In recent decades, workers have been increasingly required to cope with multiple tasks, often under strict time constraints (Howard, 1995). While in many contexts, workers' workload is centrally planned, there are a number of jobs where employees have full discretion over both the ordering and timing of each task's execution (Lyness et al., 2012). Scheduling decisions are most likely made, for example, by the so-called "knowledge workers." Engineers, physicians, architects, public accountants, lawyers, judges, and academics typically have substantial discretion on how to allocate effort across their duties. Several experts in the field believe that many more jobs in the future will be characterized by this type of scheduling autonomy (Chirkov et al., 2010; Monahan, 2018; Morgan, 2014; Pink, 2011).

This scheduling flexibility requires workers to autonomously plan and structure their work execution. Therefore, a higher degree of autonomy demands additional effort and renders individual abilities more important for successful job completion. While economists have traditionally assumed that individuals are rational in solving this effort-allocation problem, recent quasi-experimental evidence suggests that workers adopt severely inefficient work schedules (Bray et al., 2016; Coviello et al., 2014, 2015). As the authors note, this evidence might be attributed to several reasons that are both internal and external to the worker. For example, a worker may make inefficient scheduling decisions, be interrupted by coworkers, or face low incentives to exert effort. This paper experimentally tests the importance of cognitive ability in explaining scheduling choices and performance, while directly controlling for external interruptions and incentives.

From the empirical literature, very little is known about how high- and low-cognitive ability individuals decide to allocate their effort among several working activities, or on the causal effects of granting such scheduling flexibility.¹ This is not surprising since measuring the effects of scheduling flexibility on performance in the field or using observational data is challenging: task performance is not easily observed, workers could self-select into different types of jobs, or into several positions within the same job type, and employees may not have the proper

¹For experimental evidence on how individuals schedule their effort between work and on-the-job leisure activities, i.e., cyberloafing, see Corgnet et al. (2015a, 2018, 2015b,c). The works by Falk and Kosfeld (2006) and Bartling et al. (2012) examine the effect of control over minimum effort using a hypothetical principal-agent experimental framework. My setting departs from these works since I analyze how autonomy over scheduling decisions affects productivity in a real-effort framework.

incentives to exert effort. This paper seeks to fill this gap by running an experiment with high- and low-cognitive ability subjects where task performance in a cognitive-effort domain is easily observed, scheduling flexibility is randomized, and workers are equally incentivized to exert effort. In particular, this paper answers four research questions: How do workers with different cognitive ability decide to organize their workload when they have full scheduling autonomy? What are the effects of restricting this flexibility on their performance? Is scheduling ability a separate determinant of performance on top of cognitive ability? Are the results explained by time-allocation and task-switching decisions? Answers to these questions have general implications for at least two reasons especially among knowledge workers considering the cognitive nature of the tasks in the workload.² First, heterogeneity in scheduling decisions affects how these factors “aggregate up” in a specific job design. Therefore, knowing the extent of this heterogeneity becomes relevant for identifying policies designed to change the aggregate effects. For example, if low-cognitive ability workers do not manage their task load properly, they might also end up completing fewer tasks in an hour or a day of work, besides being less accurate than high-cognitive ability workers in the job performance. This is especially important in those settings where it is not possible (or too costly) to change the levels of inputs, i.e., labor and capital, to increase output. Moreover, from a theoretical perspective, it is essential to document whether cognitive ability matters in explaining scheduling decisions and, most important, to measure the behavioral biases arising in this effort allocation problem.

To address these questions, I designed a real-effort experiment with both high- and low-cognitive ability subjects who are incentivized to complete a series of cognitive tasks under different treatment conditions, which manipulates the type and extent of workers’ scheduling flexibility. Note that the high and the low types are identified using the answer they provide to the Cognitive Reflection Test – CRT ([Frederick, 2005](#)), which is a standard measure of cognitive ability.³ Moreover, considering the cognitive nature of the tasks in the workload, the so-defined

²In [Davenport \(2005\)](#) “knowledge workers” are defined as workers who can process complex or moderately complex information in mathematical and verbal form to produce knowledge (products and services). Since the real-effort tasks used in this paper demand cognitive processing such as verbal and quantitative reasoning and critical thinking the results of this paper generate more direct implications for knowledge workers than for manual labor or blue-collar personnel.

³The grouping of subjects into the high- or low- categories was done following [Frederick \(2005\)](#); namely, all those who provided 3 wrong answers to the CRT are identified as being low-ability. Note that, as suggested in [Toplak et al. \(2011\)](#), the CRT measures attributes related to cognitive ability that go beyond those measured by cognitive tests. While cognitive tests measure the computational power that is available to the participant, the CRT assesses the human tendency toward miserly processing. Additionally, it has the advantage of being a performance

high-ability workers are likely to be more capable at completing these tasks. Hence, the results of this paper fit more those labor market occupations where cognitive ability is a good proxy for skillfulness, as, for example, for knowledge workers.⁴

The experiment is divided into three parts. In the first part, subjects can freely decide which task they want to execute first, and for how long, from those in the queue. Therefore, in this part, subjects have full discretion over both the task-ordering and the time-allocation dimensions, i.e., the *Unconstrained* part. After this part is completed, participants are asked to work on another series of tasks under three treatment conditions. In one condition, they still have full discretion on how to complete this second line of work, i.e., the *Unconstrained* condition. Under the *Fixed Sequence* treatment, any arbitrary task ordering is prevented; that is, subjects cannot freely switch between tasks and instead must follow a given sequence of work (task 1, then task 2, then task 3, etc.). Under the *Fixed Time* treatment, subjects face an externally given deadline for the completion of each task. Therefore, under this condition, discretionary time allocation is restricted. Through a difference-in-differences (DID) approach that compares the within-subject performance obtained under the three treatments, i.e., *Unconstrained*, *Fixed Sequence* and *Fixed Time*, with the output obtained in the first unconstrained part of the experiment, this paper causally identifies the effects of granting different types of flexibility in a multiple-task scheduling framework. Notice that this within-subjects design helps to control for both the observed and unobserved characteristics of the individuals when estimating the treatment effects. Therefore, since it is reasonable to expect a difference in overall performance between high- and low-cognitive ability subjects, this design choice makes it possible to identify the treatment effects within each skill type. Additionally, by analyzing the task ordering and the time allocation data, this paper shows how people with different cognitive skills behave when they have full discretion or when they have limited autonomy.

The experiment is run with students enrolled in the final year of an Italian high school and the real-effort tasks are taken from a standardized test. This choice was made because it offers several advantages. First, since the real-effort tasks are taken from a test that measures performance in tasks of a general academic nature, by recruiting these students it is possible to reduce

measure rather than a self-reported variable. See [Brañas-Garza et al. \(2019\)](#) for a meta-analysis on the correlation between the CRT and decision making in economics.

⁴In the rest of paper, I defined as high- and low-skills subjects those with high and low cognitive ability, respectively.

as much as possible the dichotomy between laboratory behavior and naturally occurring decisions (Alubaydli and List, 2015; Levitt and List, 2007; Zizzo, 2010). Moreover, this framework allows one to measure skills typically demanded by the 21st-century jobs (Peterson et al., 1999) – such as verbal and quantitative reasoning and critical thinking skills – possibly drawing conclusions that go beyond the academic environment. In addition, these tasks facilitate the use of the *mouse-click tracking* technique that directly measures the individual scheduling decisions and identifies how task-ordering and time-allocation decisions explain the results. Finally, while this sample may impose some restrictions on the interpretation of the results, it also makes it possible to draw important conclusions on specific domains. First, in many contexts – such as students taking standardized tests or employee assessments – what is typically valued is a combination of ability and task-management skills; therefore, knowing the importance of each of these dimensions allows one to better derive a performance evaluation. Second, the type of task-management choices observed in this experimental framework is quite comparable to that observed in remote working jobs and to the work in the gig economy, which are becoming increasingly important, especially among young workers (Brynjolfsson et al., 2020; Pinedo Caro et al., 2021). In these types of job, individuals have indeed control over the work pace, they work on defined projects and under relative little communication (Olson, 1983). Moreover, their work is also typically built on individual, flexible and computer-based tasks (Friedman, 2014).

The aggregate effects show that imposing a deadline on the completion of each task is significantly detrimental in terms of overall performance, i.e., corresponding to a 25% reduction of the baseline mean. Whereas limits on task-ordering decisions have no effect on performance. However, the scheduling data show that individuals exhibit significant heterogeneity in their effort-allocation strategy: subjects with low cognitive ability tend to switch more across tasks and to spend more time on each task at the beginning of their time budget than high-skill subjects. Therefore, when the low-skill types are constrained in their scheduling choices, they significantly benefit from the restrictions. In particular, by preventing any switching, it is possible to significantly improve their performance by 18% of their baseline mean. On the other hand, these interventions are significantly detrimental for the high-skill subjects. Indeed, this type of worker performs better when having full scheduling autonomy.

These results can be explained by the fact that, on the one hand, the *Fixed Sequence* treatment

mechanically prevents frequent task switching, but, on the other hand, it nullifies the possibility to recognize improvement opportunities in the task ordering, which are typically not visible to the central planner. By examining the individual data on the task ordering and time allocation decisions, this paper indeed finds that low-skill subjects benefit from the *Fixed Sequence* treatment because they are prevented from switching repeatedly across tasks, allowing them to save some extra time for later tasks. On the other hand, high-skill subjects are disadvantaged by not having the possibility of postponing certain tasks, so they ultimately spend more time on initial tasks at the expense of later tasks.

Similarly, the *Fixed Time* treatment by design mechanically bounds the time possibly “wasted” on each task, but it also prevents any subjective time smoothing. The data show that high-skill individuals switch more across tasks under this treatment. Therefore, the negative effect of this treatment on these subjects is certainly induced by the constraints it imposes on the allocation of time, but also by the simultaneous change in the task switching behavior.

The paper proceeds as follows: Section 2 summarizes the related literature. Section 3 describes the experimental framework, design and procedures. Section 4 introduces the results and illustrates the robustness checks performed. Finally, Section 5 offers concluding remarks.

2 Related Literature

A long line of research has directly measured decision making in economically relevant contexts.⁵ However, direct evidence of how individuals decide to allocate their effort among multiple working activities is still lacking. In the following, I discuss how this paper contributes to the related literature, specifically focusing on the relationship between task-ordering and time-allocation decisions and performance.

Task-ordering decisions have been analyzed by the literature on multitasking.⁶ While pre-

⁵Some examples are found in games of sequential bargaining (Camerer et al., 1993; Johnson et al., 2002), in normal-form games (Costa-Gomes et al., 2001), in two-person guessing games (Gregg and Crawford, 2006), in consumption decisions over complex goods (Gabaix et al., 2006), in product purchasing (Chandon et al., 2006; Dawling et al., 2011) and in financial decisions (Duclos, 2015; Hüsler and Wirth, 2014).

⁶In the previous literature the term “multitasking” has been used to describe very different paradigms. The traditional view in this literature is to examine situations where people have to complete several tasks at once, i.e. *concurrent multitasking*, (Monsell and Driver, 2000). However, many real-world cases of task scheduling are more closely aligned with the definition of *serial multitasking* (Burgess, 2015). According to the latter definition, multitasking occurs when individuals have to complete several tasks within a certain time period, but only one task can be attempted at any time (Burgess, 2015).

vious research has widely documented the costs of task switching induced by external interruptions, the existing literature is not conclusive on the performance implications of voluntary task switching in a multiple-task framework, and it remains silent on the mechanisms behind this relationship.⁷ The paper by [Adler and Benbunan-Fich \(2012\)](#) experimentally shows that multitasking negatively affects accuracy when task switching is self-initiated. [Coviello et al. \(2015\)](#) show in a quasi-experiment that juggling more cases simultaneously slows the average case completion for Italian judges. However, no effects on accuracy emerge. In contrast, in [Buser and Peter \(2012\)](#), subjects perform significantly better when they are forced to solve two tasks sequentially than when they can freely choose the task order or when multitasking is exogenously induced. This paper builds on previous experimental studies by [Adler and Benbunan-Fich \(2012\)](#) and [Buser and Peter \(2012\)](#), and it evaluates many important new facets. First, this study analyzes the effects of preventing any task-ordering decisions while allowing for flexible time allocation. This design choice is made to test whether the previous findings also emerge in those frameworks where people can still choose how to allocate their time when they have to work sequentially. Indeed, one of the positive aspects often attributed to multitasking in the conventional wisdom is that frequent task switching is perceived as a time-saving strategy. Therefore, by not imposing any constraint on subjects' time allocation, this paper tests the performance implications of voluntary multitasking more broadly.

Another important extension relies on the collection of the task-ordering and time allocation data and on the relation of these with workers' cognitive types. In particular, by completely mapping the effort allocation strategies adopted under each experimental condition by each type of subject, this paper provides a more precise answer to the question of why multitasking affects individual performance. Finally, the real-effort tasks used in this experiment are cognitive tasks that cover several domains, i.e., verbal and mathematical reasoning and logical thinking. The first advantage of using these types of tasks, instead of, for example, Sudoku or word search puzzles, is to better approximate the type of effort demanded in many workplaces, especially among knowledge workers.⁸ Additionally, by comparing the performance obtained

⁷See [Rogers and Monsell \(1995\)](#), [Monsell \(2003\)](#) and [Foroughi et al. \(2014\)](#) for evidence on external interruptions. Although voluntary task switching has been less studied, [González and Mark \(2005\)](#) show that, among thirty-six information workers, half of the switches are due to discretionary decisions rather than external interruptions.

⁸As shown by the data from the Occupational Information Network ([Peterson et al., 1999](#)), these cognitive abilities are among the most common types of skills demanded by 21st-century jobs.

in these cognitive tasks under each experimental condition and by each type of subject, this paper further contributes to the debate on whether scheduling skills are separate from general cognitive abilities.

In this effort-allocation problem, not only task-ordering but also time-allocation decisions become crucial for production, and previous studies document that individuals do not manage their time optimally in several circumstances.⁹ Different interventions have been proposed to mitigate the consequences of this unsuccessful self-regulation. Most of the evidence is related to training interventions, but the effectiveness of this approach is still not strongly established.¹⁰ Other studies have instead used deadlines as a way to externally organize subjects' time allocation. This type of intervention indeed reduces procrastination during the completion of long-term tasks (Ariely and Wertenbroch, 2002; Bisin and Hyndman, 2014), and it also prevents suboptimal time allocation in short-term activities, including perceptual and food-choice tasks (Oud et al., 2016). However, deadlines may also induce higher time pressure, generating significant costs in terms of performance (De Paola and Gioia, 2016; Kocher and Sutter, 2006). This paper extends the above literature by studying the effectiveness of this intervention in preventing suboptimal time allocation in a multiple-task effort allocation framework and by explaining the mechanisms behind the results.

3 The Experimental Design

3.1 The Framework

The experiment uses real-effort tasks covering three cognitive domains: logical thinking and verbal and mathematical reasoning. The tasks are taken from a standardized test used by the Interuniversity Consortium (CISIA), which supports public universities in Italy in assessing students' ability and knowledge before enrollment. The standardized test is called "TOLC".

⁹Based on this writing, the ability to manage time includes "behaviors aiming at achieving an effective use of time while performing certain goal-directed activities" (Claessens et al., 2007). In the review by Claessens et al. (2007) and more recently by Aeon and Aguinis (2017), the authors summarize some of the studies investigating the problem of time management across different disciplines: sociology, psychology, management and behavioral economics. Although the results regarding how to reduce suboptimal time allocation are not yet conclusive, it is certain that humans do not yet manage this scarce resource effectively in several domains.

¹⁰Time management training typically teaches participants how to prioritize tasks, identify "time wasters", and find ways to deal with these. Despite its popularity, the evidence of the effects of such training remains mixed. While some studies show positive effects on time management behavior (Eerde, 2003; Green and Skinner, 2005), others find null effects (Häfner and Stock, 2010; Macan, 1994, 1996).

The logic domain tests the concentration aptitudes and ability to infer conclusions from precise statements. In the verbal reasoning domain, subjects have to read a brief text and then answer some comprehension questions. The mathematical tasks cover standard geometric and algebraic topics. In total, there are 34 tasks, presented in a multiple-choice format – with only one correct solution – and subjects have one hour and thirty minutes to complete them. This experimental framework offers the advantage of using real-effort tasks that require cognitive skills typically demanded by 21st-century jobs (Peterson et al., 1999), reducing as much as possible the dichotomy between laboratory behavior and naturally occurring decisions (Alubaydli and List, 2015; Levitt and List, 2007; Zizzo, 2010). Furthermore, the format of these tasks facilitates the use of the *mouse-click* tracking technique to directly measure scheduling decisions.¹¹ Thanks to this technique, it is possible not only to completely map the effort-allocation strategies adopted under each experimental condition but also to trace, at any moment in time, the exact completion rate for each task – never attempted, attempted, or completed – generating data on different dimensions of performance. For example, this technique also allows the construction of measures on the time spent solving specific tasks, on the adopted sequence of work, or on the number of times each subject has looked at the same task before completing it.

The experiment is programmed using z-Tree (Fischbacher, 2007) and follows standard experimental procedures. In particular, at the start of the experiment, subjects enter a computer laboratory where they have an isolated workstation with a computer and a connected mouse. At the start of the experiment, the software records both the time and the location of each mouse-click in the database. Figure 1 shows a screen-shot of the terminal during the experimental session. In the top part of the terminal, there are the three buttons related to the three cognitive domains; in the central part, there is the text of the first task with the associated multiple-choice options; in the bottom-left part, there are the buttons corresponding to the tasks of that specific domain; on the right of the screen, there is a time bar, which scrolls down as time passes.

Subjects participating in the experiment earn a flat payment (show-up fee) and a variable payment (piece rate), which depends on the number of completed tasks. Specifically, subjects receive 8 euros as a show-up fee and a piece rate of 1 point for each correctly completed task,

¹¹This method was first used by Payne et al. (1993) and has now been accredited by a broad set of scientific domains; see for example Camerer et al. (1993), Johnson et al. (2002), Costa-Gomes et al. (2001), Gregg and Crawford (2006) and Gabaix et al. (2006).

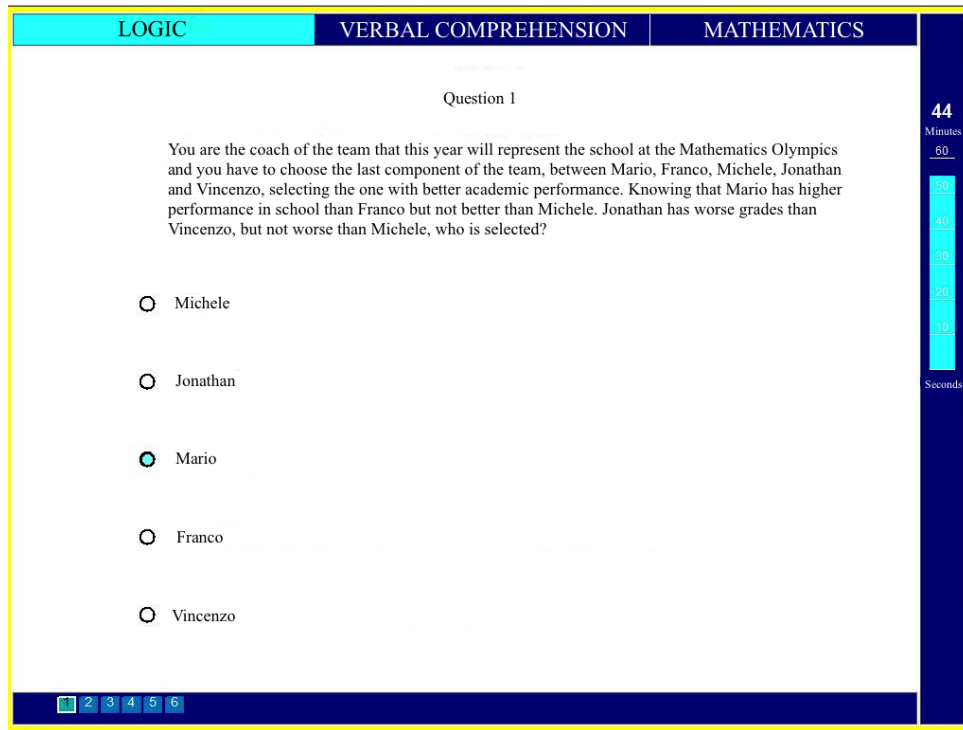


Figure 1: Ztree Screen

0 points for non-completed tasks and -0.25 points for incorrectly completed tasks. Notice that the penalty for mistakes is introduced to make random guessing unprofitable.¹² The final score is expressed in points, which are the Experimental Currency Units (ECU), and the conversion rate is 2 ECU = 1 euro. This rate implies that the maximum payment is 25 euros while the minimum is 3.75 euros.¹³

3.2 Treatments

The experiment consists of three parts. In the first part, all subjects face half of the total workload, i.e., 17 tasks – within half of the available time, i.e., 45 minutes, where six tasks are on logical thinking, five tasks are on verbal reasoning and six tasks are on mathematical thinking. In this part, subjects have full discretion on how to proceed in completing their total workload. Precisely, subjects do not face any constraint in the work sequence or in the time

¹²In addition, the regression analysis will either include individual fixed effects or controls, which also incorporate an individual measure of risk preference, so that the results should not be affected by the effect of negative marking on subjects' behavior.

¹³These amounts also include the 8 euro show-up fee. Note that the fee is chosen to more than cover the eventual loss from incorrectly completing all 34 tasks.

allocation. For the sake of simplicity, this first part is defined as *Unconstrained*.

After this first part is completed, subjects are randomly allocated to complete the second line of work under three conditions: *Unconstrained*, *Fixed Sequence* and *Fixed Time*. Under the first condition, subjects complete other 17 tasks – in the same format as the first part – and with full scheduling flexibility. Under the *Fixed Sequence* treatment, subjects must complete the remaining 17 tasks following an imposed sequence of work. Namely, they must proceed following the ascending numeration of the tasks (task 1, then 2, then 3, etc.). Therefore, if they switch tasks, they do not have the possibility to see the previous task again (the related button disappears from the screen). In the *Fixed Time* treatment, participants face a given deadline for the completion of each task, which is set by dividing the total available time by the number of tasks, i.e., $40 \text{ minutes} / 17 \text{ tasks} = 2 \text{ minutes and } 22 \text{ seconds}$.¹⁴ Since the aim of the *Fixed Time* treatment is to identify the performance consequences of limiting discretionary time allocation when task switching is still possible, as soon as the subject changes tasks, the timer for the switched task stops, and it resumes after she returns to the task. Additionally, note that in both parts and under each experimental condition, the tasks are randomly enumerated within each cognitive domain – logical thinking, verbal and mathematical reasoning – so they are not sorted on any particular dimension.

Finally, in the third part of the experiment, subjects answer a general questionnaire that collects information on sex, age, and risk-preferences – and on cognitive ability measured using the CRT by Frederick (2005) (see Section 8.3 of Appendix B). Before payment, subjects must also state their beliefs on their expected level of performance in both parts of the experiment.¹⁵ After completing the questionnaire, subjects are anonymously paid, and they leave the experimental session.

The design is summarized in the following table.

3.3 Procedures

The subject pool is composed of students enrolled in the final year of an Italian high school located in Bologna, named “Liceo Classico Statale Marco Minghetti”. According to the head

¹⁴Note that in this calculation, five minutes are left out since this was the maximum time available for reading the piece of text in the verbal reasoning part. This upper bound is given by the registered maximum value from the pilot session.

¹⁵In particular, participants are asked to state under which condition they believed they have performed better.

Table 1: Design

	Baseline	Fixed Sequence Treatment	Fixed Time Treatment
Part 1	Unconstrained	Unconstrained	Unconstrained
Part 2	Unconstrained	Fixed Sequence	Fixed Time
Part 3	Questionnaire	Questionnaire	Questionnaire

of the school, all the students enrolled in October 2015 were allowed to participate, and all the recruited subjects were at least 18 years old. The experiment took place at the BLESS Laboratory – Bologna Laboratory for Experiments in Social Science – of the Department of Economics of the University of Bologna. The recruitment was conducted in two phases. First, all the students had to complete a form to provide their contact information (personal e-mail, address of residence, mobile and home telephone numbers) and to state their preferred means of communication for receiving the participation details. In the second phase, students were randomly allocated to treatments, and they received an invitation according to the preferred means of communication, and for only specific sessions. Note that the solutions of the tasks used in the experiment were not communicated to avoid any contamination across sessions. Moreover, the results would be interpreted after controlling for within-session correlation.

The experiment started with an introduction explaining the instructions related to the first part and with a practice example to let the students become familiar with the programmed Ztree interface. The instructions related to the second part of the experiment circulated only after every subject completed the first part. In total, 87 subjects participated, and their performance was observed over 34 tasks; therefore, the results in the next section will be interpreted using panel data estimation techniques.¹⁶ Overall, the experiment lasted 1 hour and 45 minutes and the average payment was approximately 15.30 euros.

¹⁶Longitudinal study designs are useful to investigate changes in the outcome over time within subjects and to compare these changes among treatment groups. From a statistical perspective, panel data usually allow the researcher to increase the precision of the estimates, thus increasing the power to detect treatment effects. Indeed, between-subjects design requires from 4 to 8 times more participants than a within-subject design to reach an acceptable level of statistical power [Bellemare et al. \(2016\)](#).

4 Results

This section first investigates subjects scheduling decisions during the first unconstrained part of the experiment, particularly focusing on differences in behavior between high- and low-skilled subjects. The final part of the section presents the treatment effects and the related mechanisms, and it describes the robustness of the results.

4.1 Scheduling Decisions

In the first part of the experiment, subjects are free to organize their workload. In particular, they can make autonomous decisions over both the task-ordering and the time-allocation dimensions. Since this level of autonomy demands additional effort, thus conferring greater importance on individual abilities for successful job completion, I first examine whether cognitive ability matters in explaining subjects' effort allocation strategies and performance.

As a first step, I examine how the answers to the CRT ([Frederick, 2005](#)) are distributed in the sample. Indeed, since this test captures the ability to focus and engage in deep reflection over the tendency to implement fast and frugal judgments, the CRT is often used as a standard measure of a specific type of cognitive ability ([Frederick, 2005](#)). In particular, the answers to the CRT are grouped into a single indicator variable, which is equal to 1 if the subject provided a wrong answer to all the three questions of the test ([Frederick, 2005](#)). Table A4 of Appendix A shows the distribution of subjects into this categorical variable. Note that more than 50% of the participants are identified as being low skilled. Note that the distribution of answers to the CRT is very similar to previous studies; for a review, see [Brañas-Garza et al. \(2019\)](#).

The first dimension of the effort allocation problem considered here is the task ordering. Previous literature has shown that voluntary task switching might decrease performance ([Buser and Peter, 2012](#)). However, it is also possible that having the flexibility of self-ordering the tasks could increase workers' productivity since subjects may recognize improvement opportunities that are not visible to central planners ([Tucker, 2007](#)). Unless otherwise specified, task-switching behavior is defined by counting the number of times each subject looks at the same task before completing it, i.e., *lookups*. Table A3 of Appendix A provides summary information on this measure and on other sample statistics by cognitive types. Note that both high- and low-skill subjects looked on average nearly 3 times at each task and that there is substantial

heterogeneity in this behavior.¹⁷ However, if we examine the distribution of lookups over time by each skill-type in Figures A1 and A2, we observe that while high-skilled workers return to previous tasks mostly at the end of their total time, low-skilled subjects look at previous tasks – before completing the others in the queue – at the beginning. In other words, it seems that low-skilled subjects “get stuck” on certain tasks, even if, by completing other tasks, they can obtain a higher payoff. This difference in the task-ordering behavior between high- and low-skilled subjects is confirmed through regression analysis. The first column of Table 2 shows that while the number of lookups increases significantly over time for the high-skill group, the opposite is true for the low-skill types who look less often at previous tasks toward the end of their time (the coefficient t_2 is negative and significant in Column 1 and 3).

The second dimension of this scheduling problem involves the allocation of time. Previous studies have shown that in food-choice and perceptual tasks, people tend to spend too much time on options with a lower reward with respect to the other alternatives (Oud et al., 2016). The current experimental framework offers a constant piece rate for the completion of each task. Therefore, the payoff of an “easy” task is the same as the returns from a “difficult” task. A direct implication of this incentive scheme is that subjects should proceed by postponing those tasks that are dominated by easier options. The summary statistics of Table A3 of Appendix A show that low-skill subjects spend more time on tasks than high-skill workers on average, especially on the completed tasks. If we consider Figure A3 and the second column of Table 2, we can observe that, for high-skilled subjects, the time spent on each task increases over time, and this is driven by the fact that they look at difficult tasks mostly at the end – the time spent only on completed tasks, in column six, is lower than the time spent on all tasks in column four of Table 2. Low-skilled workers instead spend a higher share of their total time budget at the beginning, significantly exhausting the available time to complete later tasks; see the coefficients in columns two and four at the bottom of Table 2.

Taken together, this evidence suggests that when subjects have to solve this unconstrained effort-allocation problem, high- and low-skilled workers make significantly different scheduling decisions. In particular, high-skilled workers do not spend much of their total time on tasks that are dominated by easier options – especially at the beginning – and they look back at

¹⁷This is line with previous evidence showing that in a 6-task framework, switches ranged from 5 to 29 (Adler and Benbunan-Fich, 2012) and that in a 2-task scenario, the mean number of switches is 2.16 with a standard deviation of 2.20 (Buser and Peter, 2012).

Table 2: Behavior in the First Part by High/Low Skill Types

	All				Completed	
	Lookups	Time on Task	Lookups	Time on Task	Lookups	Time on Task
<i>High Skill</i>						
t_2	0.855*** (0.180)	2.201 (8.345)	0.642** (0.244)	13.000 (10.460)	0.670** (0.262)	4.889 (12.672)
t_3	2.203*** (0.182)	53.937*** (8.143)	2.115*** (0.268)	78.193*** (10.430)	2.037*** (0.320)	67.112*** (14.723)
<i>Low Skill</i>						
t_1	0.311 (0.188)	34.670*** (9.174)				
t_2	-0.443* (0.260)	-29.459*** (10.756)	-0.477* (0.273)	-39.242*** (12.215)	-0.329 (0.279)	-21.889 (16.284)
t_3	-0.298 (0.238)	-35.192*** (12.436)	-0.426 (0.328)	-50.188*** (15.407)	-0.344 (0.394)	-33.891 (21.865)
Constant	5.826* (3.211)	-35.900 (52.240)	2.903*** (0.120)	107.733*** (6.135)	2.890*** (0.147)	96.433*** (7.809)
Observations	1479	1479	1479	1479	953	953
Time fixed effect	✓	✓	✓	✓	✓	✓
Section fixed effects	✓	✓	✓	✓	✓	✓
Clustered Std. Errors	✓	✓	✓	✓	✓	✓
Individual fixed effects			✓	✓	✓	✓
Adj.R ₂	0.293	0.046	0.359	0.012	0.398	0.010

Notes: *Lookups* and *Time on Task* are the outcome variables. *Lookups* counts the number of times the task was seen by the same subject before completion. *Time on Task* measures the seconds spent on each task. *Low Skilled* is a dummy variable equal to one if the subject wrongly answered all the 3 questions of the CRT. The regression results are estimated using the following model:

$$Y_{itjs} = \alpha + \beta_1 \text{Low Skill} + \beta_2 \text{Low Skill} \times T + \delta_i T + \gamma_i + \theta_s + \epsilon_{itjs} \quad (1)$$

where the subscript i indicates the individual, t the time, j the task and s the section, and the parameters γ_i , δ_t , θ_s account for the individual, time and section fixed effects, respectively. The coefficients t_1 - t_3 under the Low Skill panel shows the interactions of time fixed effects with the Low-skill dummy. Notice that the time fixed effects t_1 - t_3 are constructed so that each time effect represents a range of time of 15 minutes. The last two columns look at completed tasks only. Errors are clustered at the individual and session levels. Standard errors reported in parentheses, and *, **, and *** indicate statistical significance at the 10%, the 5%, and the 1% levels, respectively.

previous tasks only toward the end of their time. In contrast, low-skilled subjects tend to “get stuck” on certain tasks at the beginning of their total time, even if, by completing other tasks, they can obtain a higher payoff.

To further understand whether these differences in scheduling strategies generate a difference in performance, Table A6 shows how the probability of completing one task changes between low- and high-skilled subjects. In particular, the first column of Table A6 shows that there is a significant baseline difference in performance between these types of workers by around 8 percentage points when controlling for individual observable characteristics. This gap in performance persists even when time and section fixed effects are included, as the dummy variable *Low Skill* is negative and statistically significant (Columns 2 and 3). This indicated that, in the first 15 minutes, low-skill subjects have a lower probability of completing a task by around 14 percentage points with respect to the high-skill (i.e the Constant term). However, when we control for observable task-ordering and time-allocation decisions, there is no residual gap in the performance between low- and high-skilled subjects, as the evolution of performance over time of the low-skilled does not significantly differ from that of the high-skilled workers – neither the dummy *Low Skill* or the time fixed effect t_2 and t_3 in the “Low Skill” panel are significant. This suggests that the heterogeneity in the scheduling behavior explains most of the variation in outcomes between these types of workers, in the first part of the experiment.¹⁸

4.1.1 Treatment Effects

As shown in Table 1, in the second part of the experiment, subjects are randomly allocated to three treatments. Summary statistics and balance tests are presented in Tables A1 and A2 of the Appendix. As these tables show, no significant differences emerge in any of the observable characteristics across treatment groups.¹⁹

Unless otherwise specified, the results on treatment effects are presented considering as

¹⁸When individual fixed effects, which control for both the observed and unobserved characteristics of the individual are also included, there is no residual gap in performance. These results are included in the <https://dx.doi.org/10.2139/ssrn.4259226>

¹⁹Note that the sample size was calibrated to reject with an 80% power in a longitudinal study design with 17 repetitions per subject per part, a treatment effect from 13% to 15% of the mean value of the control group, using an average value of 0.65 for the probability to complete one task and a standard deviation of 0.33 in the control group. Note that while the effect size was taken from the results of Buser and Peter (2012), both the control group values on the average and on the standard deviation were taken from the pilot data.

the output of interest the probability of completing one task and using the following panel regression model:

$$Y_{itjs} = \alpha + \beta_1 \text{Part} + \beta_2 \text{Treatment} + \beta_3 \text{Part} \times \text{Treatment} + \gamma_i + \delta_t + \theta_s + \epsilon_{itjs} \quad (2)$$

where the subscript i indicates the individual, t the time, j the task and s the section, and the parameters γ_i , δ_t , θ_s account for the individual, time and section fixed effects, respectively. The coefficient of interest is β_3 , which identifies the difference across parts and between treatment groups, i.e. the Difference-In-Differences (DID) estimator. Section 4.3 tests the robustness of the results to different outcome measures. Table 3 shows how the average probability of completing one task changes across parts and treatment groups, i.e., the aggregate treatment effects.

Table 3: Average Treatment Effects

	(1)	(2)	(3)	(4)
Fixed Sequence	-0.024 (0.029)	-0.028 (0.029)		
Fixed Time	-0.032 (0.029)	-0.038 (0.029)		
Second Part x Unconstrained	0.027 (0.028)	0.028 (0.028)	0.029 (0.028)	0.029 (0.034)
Second Part x Fixed Sequence	-0.006 (0.041)	-0.036 (0.041)	-0.038 (0.040)	-0.038 (0.043)
Second Part x Fixed Time	-0.159*** (0.041)	-0.162*** (0.041)	-0.162*** (0.040)	-0.162*** (0.043)
Constant	0.663*** (0.020)	0.713*** (0.028)	0.693*** (0.024)	0.693*** (0.020)
Observations	2958	2958	2958	2958
Section fixed effects	✓	✓	✓	✓
Time fixed effects		✓	✓	✓
Individual fixed effects			✓	✓
Clustered Errors				✓
Adj. R^2	0.090	0.097	0.143	0.142

Notes: $P(\text{Complete a Task})$ is the outcome variable, which measures the average probability of completing one task correctly. *Fixed Time* and *Fixed Sequence* are the dummy variables indicating the treatment groups. *Second Part* is a dummy variable, which equals one if the second part of the experiment is considered. Errors are clustered at the individual and session levels. Standard errors reported in parentheses, and *, **, and *** indicate statistical significance at the 10%, the 5%, and the 1% level, respectively.

Table 3 shows that there are no statistically significant differences in performance between each treatment group in the first part of the experiment. In the second part, subjects under

the Fixed Sequence treatment have to work on the tasks without being able to self-order them, while subjects in the Unconstrained group still have full flexibility in choosing how to organize their workload. The results show that even in this second part, the two groups achieve similar performance, suggesting that, on average, limits on task switching generate an output similar to the flexible scheme, i.e., the coefficient *Second Part x Fixed Sequence* is not statistically significant.

However, Table 3 shows that imposing a given deadline per task (2 minutes and 22 seconds) significantly decreases overall output. Indeed, in the second part of the experiment, subjects under the Fixed Time treatment have a lower probability of completing a task by more than 16 percentage points, which corresponds to a decrease in performance of more than 25% of the mean value.

Therefore, on average, imposing a constraint on subjects' time allocation is significantly detrimental for overall performance. Given the observed heterogeneity in the effort allocation strategies adopted by the different types of subjects, the next section further examines how the treatment effects change between high- and low-skilled workers.

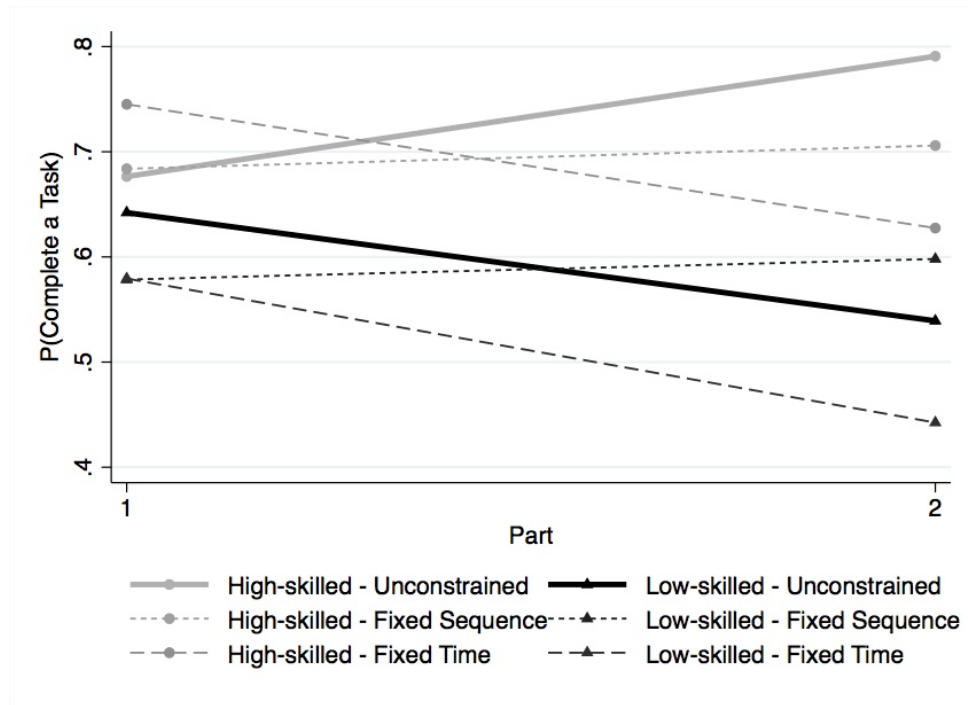
4.2 Heterogeneous Treatment Effects

This section explores whether there is any heterogeneity in the treatment effects, and it clarifies the mechanisms behind the results. From the review of the related literature, it is indeed reasonable to expect the treatment restrictions to be heterogeneous. In particular, when subjects have complete autonomy in managing their workload, they not only have to devote effort into solving each task, but they also have to plan and schedule their work execution, introducing an additional decision for successful task completion. Therefore, on the one hand, the imposed constraints might alleviate the burden of these scheduling decisions, but on the other hand, they nullify the possibility to exploit improvement opportunities in the task ordering or in the time allocation. For these reasons, it is intuitive to expect low-skill workers to benefit from having less decisional power, but high-skill workers to profit from having the possibility of self organizing.

Figure 2 graphically analyzes these hypotheses by showing how the mean probability of completing one task varies between workers with different cognitive ability across treatments

and parts of the experiment. As the figure shows, high-skill workers reach the highest performance when they have full autonomy (solid light gray line), they significantly underperform when they have a deadline per task (long-dashed light gray line), while limits on the task ordering have still a negative impact on their performance but of a lower magnitude (short-dashed light gray line). However, low-skill subjects significantly increase their performance when they face a given sequence of tasks (short-dashed dark gray line) with respect to when they are unconstrained (solid dark gray line). Additionally, even for this type of workers, deadlines are overall detrimental in terms of performance, although their output decrease to a lesser extent than among high-skill subjects.

Figure 2: Average Probability of Completing one Task across Treatments and Skill-Types



To further assess the robustness of this graphical evidence, Table 4 presents the regression results of the treatment effects on the full sample and separately for each type of worker. Table 4 confirms the results found in the graphical inspection of the data. In particular, the Fixed Sequence treatment significantly increases the performance of the low-skilled subjects to a greater extent than for high-skill workers – the coefficient of *Low Skill* \times *Second Part* \times *Fixed Sequence* is

strongly significant and greater in magnitude than the coefficient *Second Part x Fixed Sequence*, so that the overall effect of this treatment on low-skill subjects is positive and significant (Column 7). Specifically, when this type of subject does not have to consider the task ordering, it has a higher probability of completing one task by approximately 11 percentage points. In other words, under this constraint, they fill the performance gap that, otherwise, they would have experienced under the flexible job design due to poor task-management – the coefficient of *Second Part x Unconstrained* is negative and of similar magnitude in Column 7. On the other hand, when high-skill workers cannot arrange the tasks as they prefer, they underperform with respect to the unconstrained case – the coefficient of *Second Part x Fixed Sequence* reported in Column 6 is significantly negative, while the coefficient of *Second Part x Unconstrained* is positive and significant.

Table 4 also confirms the impact of the Fixed Time treatment, which restricts subjects' time allocation decisions. The results indeed show that imposing a given deadline per task decreases the performance of the high-skill subjects – the coefficient *Second Part x Fixed Time* is negative and significant in Column 6. However, while the effect of this treatment on the performance of low-skilled workers is still negative overall, the decline in performance among this type of workers is lower than that experienced by high-skill workers – the coefficient *Low Skill x Second Part x Fixed Time* is positive and strongly significant in Column 5. As shown in Table 4, the results hold under the inclusion of fixed effects and the clustering of the standard errors.²⁰

Table 4 also shows that, while the high-skill subjects significantly increase their performance over time when unconstrained – the coefficient of *Second Part x Unconstrained* in Column 6 is positive and statistically significant – the low-skill type did not – the coefficient of *Second Part x Unconstrained* in Column 7 is negative and significant. This result suggests there could exist significant differences in how these two types of subjects learn to work on the tasks or in the fatigue they experienced across parts, which is in line with the study by [Brown et al. \(2022\)](#). To better understand the sources of this difference, Table A5 in the Appendix looks at how subjects' behavior changed across parts in the unconstrained group. While, when looking

²⁰Section fixed effects control for cognitive domain general effects: they account for the fact that performance in a specific cognitive domain may be generally higher than that in others. Time fixed effects control for the general effect of time, to capture, for example, fatigue or learning. Individual fixed effects account for individual-specific observed and unobserved characteristics. The clustering of the standard errors at the individual level is instead used to account for the autocorrelation in the error term due to repeated observation within individuals.

Table 4: Heterogeneous Treatment Effects by Types

	(1) Full	(2) Full	(3) Full	(4) Full	(5) Full	(6) High Skill	(7) Low Skill
Fixed Sequence	0.007 (0.039)	0.011 (0.038)	0.008 (0.037)				
Fixed Time	0.069 (0.044)	0.066 (0.042)	0.064 (0.042)				
Low Skill	-0.034 (0.043)	-0.024 (0.041)	-0.017 (0.041)				
Low Skill x Fixed Sequence	-0.071 (0.062)	-0.075 (0.060)	-0.083 (0.059)				
Low Skill x Fixed Time	-0.131** (0.062)	-0.110* (0.060)	-0.120** (0.059)				
Low Skill x Second Part x Unconstrained	-0.217*** (0.060)	-0.217*** (0.057)	-0.231*** (0.057)	-0.231*** (0.057)	-0.231*** (0.056)		
Low Skill x Second Part x Fixed Sequence	0.215** (0.087)	0.215*** (0.083)	0.226*** (0.083)	0.227*** (0.082)	0.227*** (0.075)		
Low Skill x Second Part x Fixed Time	0.198** (0.089)	0.198** (0.085)	0.210** (0.084)	0.210** (0.084)	0.210*** (0.076)		
Second Part x Unconstrained	0.114*** (0.036)	0.114*** (0.034)	0.118*** (0.034)	0.118*** (0.033)	0.118*** (0.035)	0.118*** (0.035)	-0.113** (0.044)
Second Part x Fixed Sequence	-0.092* (0.053)	-0.092* (0.051)	-0.116** (0.051)	-0.117** (0.050)	-0.117** (0.047)	-0.116** (0.048)	0.109* (0.059)
Second Part x Fixed Time	-0.232*** (0.064)	-0.232*** (0.061)	-0.242*** (0.061)	-0.242*** (0.061)	-0.242*** (0.053)	-0.242*** (0.052)	-0.032 (0.055)
Constant	0.676*** (0.027)	0.550 (0.391)	0.616 (0.384)	0.690*** (0.020)	0.690*** (0.017)	0.741*** (0.018)	0.637*** (0.029)
Observations	2958	2958	2958	2958	2958	1462	1496
Section fixed effects		✓	✓	✓	✓	✓	✓
Time fixed effects			✓	✓	✓	✓	✓
Individual fixed effects				✓	✓	✓	✓
Clustered Errors					✓	✓	✓
Adj.R ₂	0.037	0.12	0.13	0.15	0.15	0.11	0.14

Notes: $P(\text{Complete one Task})$ is the outcome variable. *Low Skill* is a dummy variable equals to one if the subject has wrongly answered all the 3 questions of the CRT. *Fixed Time* and *Fixed Sequence* are the dummy variables indicating the treatment groups. *Second Part* is a dummy variable equal to one if the second part of the experiment is considered. Errors are clustered at the individual and session levels. Standard errors reported in parentheses, and *, **, and *** indicate statistical significance at the 10%, the 5%, and the 1% level, respectively.

at changes in the task-ordering and time allocation behavior, we do not see a significant shift across parts of the experiment for both types of subjects (Columns 1 and 2 for the high-skilled and 3 and 4 for the low-skilled), when we look at the probability of making a mistake we see significant differences. In particular, the high-ability workers tend to make significantly fewer mistakes in the second part of the experiment when unconstrained (Column 3), while lower ability subjects increase their chances of giving an incorrect answer. This suggestive evidence signals significant differences in the fatigue or in the ability to focus experienced across parts by the two cognitive types, further confirming their baseline disparity in terms of capability, which indeed might significantly exacerbates their performance gap over time in the unconstrained job design.

To better understand the mechanisms behind the results, Figures 3 - 6 use the scheduling data to clarify how the behavior of each type of subject changes under each treatment between the first and second parts of the experiment. Figures 3 and 4 show that, although the Fixed Sequence treatment is specifically designed to restrict task-ordering decisions, it has also induced a change in the time spent on each task by both types of workers. In particular, under this treatment, low-skill subjects not only do not switch across tasks, but they also reduce the time spent on initial tasks – increasing the time available for later tasks – suggesting that they benefit from this treatment by not switching repeatedly, but also by proceeding more quickly through the initial tasks. On the other hand, high-skilled subjects, by not having the possibility of postponing certain tasks, they ultimately spend more time on initial tasks at the expense of later tasks.

Figures 5 and 6 show the mechanisms behind the Fixed Time treatment. Under this treatment, which imposes strict bounds on the time spent on the completion of each task, low-skill subjects look at previous tasks more toward the end of their time. While high-skill individuals switched more across tasks at any moment in time.

This graphical evidence is also confirmed by the regression analysis in Tables 5 and 6. Columns 1 and 5 of Table 5 show how the number of lookups changes for high- and low-skilled workers under the Fixed Sequence treatment with respect to the Unconstrained baseline. As expected, the treatment mechanically reduces the number of lookups for both types of workers. As the graphical inspection of the data has shown, the Fixed Sequence treatment also induces the low-skill subjects to proceed more quickly through the initial tasks and to spend more time

Mechanism behind the Fixed Sequence Treatment

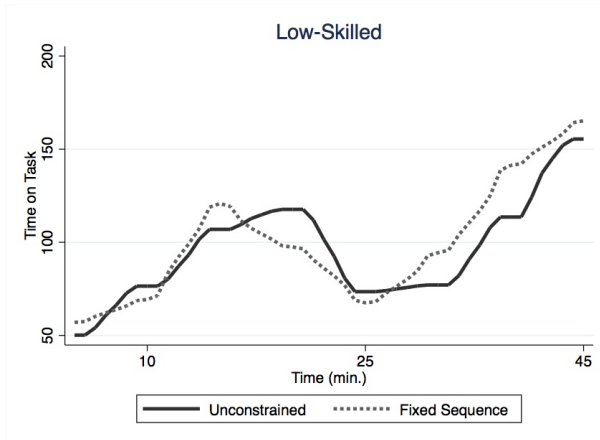


Figure 3: Low-Skilled - Time on Task

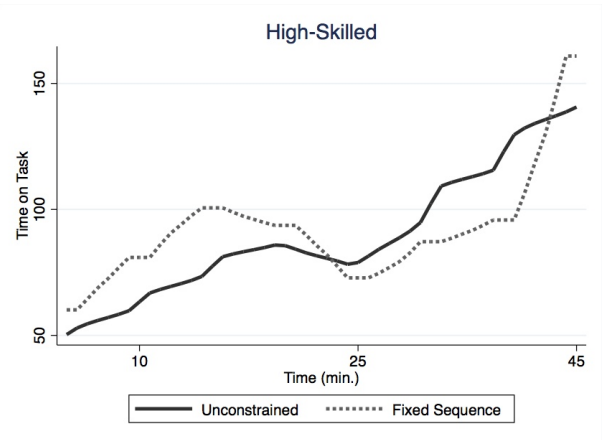


Figure 4: High-Skilled - Time on Task

Mechanism behind the Fixed Time Treatment

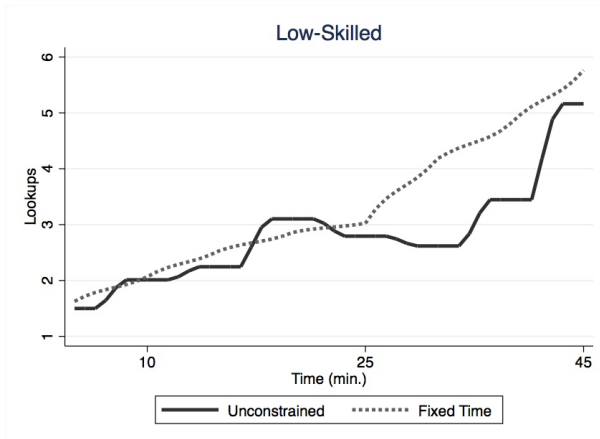


Figure 5: Low-Skilled - Lookups

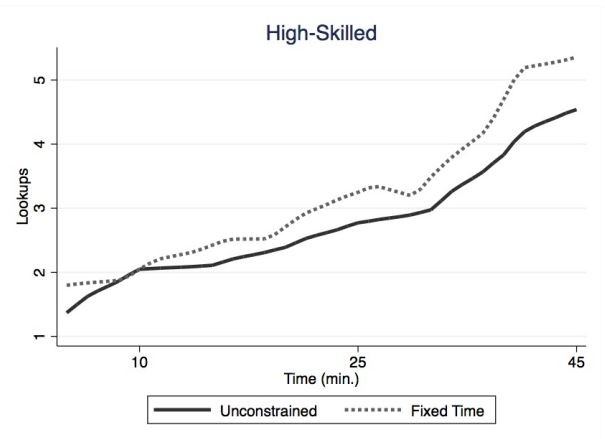


Figure 6: High-Skilled - Lookups

on later tasks – the coefficients on the later time dummies become positive under the Fixed Sequence treatment (Column 6). On the other hand, high-skill subjects under this treatment ultimately have less time to complete later tasks (Column 2).

Similarly, Table 6 shows the full regression results on the effects of the Fixed Time treatment on both the task-ordering and the time-allocation dimensions. Column 5 confirms that this treatment increased the time spent on later tasks among the low-skilled. Moreover, this treatment also changed the number of lookups done by the high-skill subjects, especially toward the end of their time (Column 4).

Table 5: Mechanism behind the Fixed Sequence Treatment by High/Low Skilled Types

	High Skill				Low Skill			
	Lookups	Time on Task	Lookups	Time on Task	Lookups	Time on Task	Lookups	Time on Task
<i>Unconstrained</i>								
t_2	0.717*** (0.126)	6.278 (6.002)	0.628* (0.311)	6.497 (4.465)	0.730* (0.312)	-49.410*** (10.736)	0.517** (0.159)	-45.708** (12.746)
t_3	2.074*** (0.316)	42.970** (11.216)	1.986** (0.508)	49.224** (15.594)	2.301*** (0.280)	20.768 (11.457)	2.121*** (0.189)	31.041** (10.989)
<i>Fixed Sequence</i>								
t_1	-0.205 (0.259)	27.890** (8.972)			-0.671*** (0.142)	-17.470*** (2.982)		
t_2	-0.343* (0.134)	-19.452 (11.449)	-0.271 (0.198)	-20.676 (13.128)	0.228 (0.283)	32.689*** (6.279)	0.343** (0.097)	30.853** (8.911)
t_3	-2.103*** (0.386)	-47.027* (20.494)	-2.103*** (0.521)	-64.904* (28.037)	-1.262* (0.530)	39.307* (18.245)	-1.296** (0.438)	34.690 (19.665)
Constant	1.487 (1.720)	32.832 (89.867)	2.552*** (0.213)	117.148*** (4.606)	14.784*** (2.004)	101.251 (72.585)	2.462*** (0.156)	123.775*** (7.547)
Observations	578	578	578	578	408	408	408	408
Time fixed effect	✓	✓	✓	✓	✓	✓	✓	✓
Section fixed effect	✓	✓	✓	✓	✓	✓	✓	✓
Clustered Std. errors	✓	✓	✓	✓	✓	✓	✓	✓
Individual fixed effects			✓	✓			✓	✓
<i>Adj.R₂</i>	0.419	0.059	0.459	0.038	0.379	0.077	0.402	0.043

Notes: *Lookups* and *Time on Task* are the outcome variables. *Lookups* counts the number of times the task was seen by the same subject before completion. *Time on Task* measures the seconds spent on each task. Time fixed effects t_1 - t_3 are constructed so that each time effect represents a range of time of 15 minutes. The regressions include individual, time and section fixed effects. Errors are clustered at the individual and session levels. Standard errors reported in parentheses, and *, **, and *** indicate statistical significance at the 10%, the 5%, and the 1% level, respectively

Finally, to test more directly whether differences in the task-ordering and time allocation decisions between skill types explain most of their performance differences, I create two behavioral measures describing subjects' effort allocation in each part, and I test whether these measures capture most of the change in performance across treatments. In particular, to sum-

Table 6: Mechanism behind the Fixed Time Treatment by High/Low Skilled Types

	High Skill				Low Skill			
	Time on Task	Lookups	Time on Task	Lookups	Time on Task	Lookups	Time on Task	Lookups
<i>Unconstrained</i>								
t_2	9.114 (13.362)	0.770** (0.337)	10.074 (14.002)	0.690* (0.386)	-24.756* (14.534)	0.916** (0.431)	-24.454 (15.804)	0.558 (0.455)
t_3	45.174*** (12.300)	2.037*** (0.256)	52.171*** (17.407)	1.945*** (0.348)	39.361*** (13.806)	2.331*** (0.473)	46.310*** (14.309)	2.055*** (0.559)
<i>Fixed Time</i>								
t_1	-15.221 (15.072)	0.228 (0.305)			-25.833* (14.555)	0.286 (0.431)		
t_2	3.444 (19.687)	-0.089 (0.515)	-0.566 (21.533)	0.246 (0.513)	48.132*** (16.755)	0.367 (0.541)	47.764** (18.839)	0.543 (0.586)
t_3	16.179 (20.139)	0.628 (0.544)	9.997 (30.558)	1.259* (0.702)	10.279 (16.373)	0.478 (0.614)	-7.298 (17.537)	0.136 (0.658)
Constant	24.221 (50.365)	1.656 (1.621)	95.745*** (8.950)	2.572*** (0.227)	175.863** (76.664)	6.307 (5.746)	105.972*** (6.320)	3.002*** (0.210)
Observations	459	459	459	459	543	543	543	543
Time fixed effect	✓	✓	✓	✓	✓	✓	✓	✓
Section fixed effect	✓	✓	✓	✓	✓	✓	✓	✓
Clustered Std. errors	✓	✓	✓	✓	✓	✓	✓	✓
Individual fixed effects			✓	✓			✓	✓
$Adj.R_2$	0.104	0.280	0.087	0.315	0.053	0.297	0.034	0.374

Notes: *Lookups* and *Time on Task* are the outcome variables. *Lookups* counts the number of times the task was seen by the same subject before completion, over time. *Time on Task* measures the seconds spent on each task, over time. Time fixed effects t_1 - t_3 are constructed so that each time effect represents a range of time of 15 minutes. The regressions include individual, time and section fixed effects. Errors are clustered at the individual and session levels. Standard errors reported in parentheses, and *, **, and *** indicate statistical significance at the 10%, the 5%, and the 1% level, respectively.

marize the task switching behavior, I created an individual measure (which varies across parts) counting the average number of times a person switches before completing a task, by taking the within-part average of the “Lookups” per task measure; while to capture time allocation efficiency – which in this context coincide with spending more and more time on later tasks – I created a variable (which varies across parts) measuring the correlation between the time spent on the task and the clock time, running simple individual-level OLS regressions of the “Time on task” measure on the clock time and forcing the slope through the origin.²¹ In this way, it is possible indeed to identify the degree of efficiency in the time allocation decisions, since a positive slope indicates that the subject has spent more and more time on later tasks. Then I estimate the model shown in Table A7 to test whether these measures capture most of the performance differences between high- and low-skilled subjects across treatment groups. In particular, the first column of Table A7 estimates the overall difference in performance between high- and low-skill subjects both in the first and the second part of the experiment, without including any of the behavioral measures. The estimates show a significant gap in performance between high- and low-skills subjects in the first part when they have full autonomy; however, when the behavioral measures are included, the performance gap reduces and it becomes not statistically significant anymore – the dummy *Low Skill* in the second column is not significant and it is lower in magnitude. Similarly, in the second part of the experiment, the low-skill subjects perform significantly worse than the high-skilled workers – the interaction term *Low Skill x Second Part* is negative and statistically significant, while the dummy *Second Part* is positive in Column 1 – however, when we account for the behavioral differences across types, no residual performance gap emerges in Column 2. This is also shown for each treatment group separately: the change in behavior induced by each treatment in the second part (Fixed Sequence in Column 6 and Fixed Time in Column 8) is capturing all the performance differences between types, as the dummy variables *Second Part*, *Low Skill* and *Low Skill x Second*

²¹The first behavioral measure, i.e., *Mean Lookups*, counts the average number of times the subject has switched task (before completion) in each part, i.e., $Mean\ Lookups = \frac{\sum_{n=1}^{17} Lookups\ per\ task_n}{17}$, where n represent the task and 17 is the total number of tasks in each part. While, the *Time Correlation* measure is given by the correlation between the time spent on each task and the time passed by, as captured by the β coefficient of the following regression (run separately for each individual within the same part):

$$Time\ on\ Task_n = \beta\ Time + \epsilon_n \quad (3)$$

where n represent the task, and *Time* is the clock time.

Part are not significant anymore. Notice however that in the Unconstrained group, there still exists a performance gap between high- and low-skill subjects, even after controlling for their different behavior across parts (Column 4). This indicates that the gap in performance between the low- and high-skill subjects observed in the Unconstrained group, could not be fully explained by their different effort allocation strategies, but possibly by their baseline disparity in terms of capability. A result, which is indeed in line with the evidence shown in Table A5 indicating that contrarily to the high-skilled, the low-skill subjects made more mistakes in the second part of the experiment.

Overall, these results suggest the types of workers who benefit the most from having full scheduling autonomy and how it is possible to increase the performance of the low-skill workers. In particular, the experiment shows that, on the one hand, by controlling the sequence of work, it is possible to eliminate the costs due to the lower task-management ability of the low-skill workers – although the gap in performance due to their lower cognitive ability still persists with respect to the high-skill subjects. Moreover, this restriction on the task-ordering dimension reduces the performance of the high-skilled, suggesting that it is important to know the share of high- and low-skill workers in the population of interest to understand if this constraint could generate an overall increase in efficiency. On the other hand, imposing a deadline per task produces significant costs on the performance of both types of workers, although the reduction is less pronounced among the low-skill subjects. To further support the validity of these results, the next section runs some robustness checks.

4.3 Robustness Checks

The first robustness check replicates the above analysis by using different performance outcomes. The first column of Table A8 shows how the estimates change when the final score is used as the main outcome of interest. This measure of performance sums up all the points obtained by each subject for the completed tasks, and it subtracts the total penalty for the wrongly completed tasks. The previous findings are confirmed. In particular, high-skilled workers perform significantly worse when they are constrained on either the task-ordering or the time-allocation dimension – the Fixed Sequence treatment reduces performance by approximately 2 points, while the Fixed Time treatment reduces the performance by approximately 4 points in

the final score. In contrast, low-skilled workers, who tend to switch repeatedly across tasks and to not prioritize their work, perform better under both treatments than when unconstrained, with respect to high-skilled workers – both the Fixed Sequence and the Fixed Time treatment increase performance by approximately 4 points, which is similar in magnitude to the *Second Part x Unconstrained* coefficient.

Columns two and three replicate the main findings of Table 4 using as main outcomes the probability of making a mistake and of not completing a task. Note that both treatments increase the performance of low-skilled workers since they induce them to make fewer mistakes and to not complete certain tasks. High-skilled workers, under both treatments make more mistakes, while there is no difference in the probability of not completing a task in the queue.

Table A9 shows that the main results are robust to the use of continuous behavioral measures as a way to identify low- and high-skill subjects, especially when they are identified by their task-ordering decisions. In particular, from the previous sections, we have seen that low-skilled subjects looked up more frequently and spent more time on difficult tasks before completing the others in the queue. Therefore, Table A9 uses the mean number of lookups made and the mean time spent on the erroneously completed tasks – given that these behavioral measures should be clearly correlated with worker type – to check the robustness of treatment effects estimated using the CRT classification. As Table A9 shows, for each additional lookup done on the incorrectly completed tasks in the first part of the experiment, the Fixed Sequence treatment increases performance by 8 percentage points in the second part. Note that, these results not only confirm previous findings but also use a higher sample variation since the continuous measures exploit the full heterogeneity across subjects, therefore increasing the statistical power of the significance tests [Altman and Royston \(2006\)](#).

An additional robustness check runs a similar analysis using logistic panel regressions. These regression results are shown in Table A10 and the marginal effects in Table A11. Both the sign and the significance of the heterogeneous treatment effects are confirmed even when the hypothesis of a linear relationship is relaxed.

5 Conclusions

This paper contributes to the literature investigating the importance of cognitive ability in effort-allocation decisions. The experiment builds on previous evidence showing that constraints on individual decision making could improve performance, and it extends the literature to a case where both high- and low-cognitive ability subjects are observed working on multiple tasks, demanding real cognitive effort, and where data on individual scheduling decisions are collected.

The data show that, on average, restricting time-allocation decisions is significantly detrimental in terms of overall performance. However, constraints on task-ordering decisions generate an outcome similar to the fully unconstrained case. Observing the worker type, however, is an important element, as the impact of these constraints changes with the cognitive ability of subjects. In particular, the low-cognitive ability workers tend to intensively switch across tasks when they have complete autonomy in scheduling their workload. Therefore, when these types of workers are constrained on the task-ordering, they significantly benefit from having less flexibility. Moreover, this centralized job design reduces the performance of high-cognitive ability workers, suggesting that it is important to know the share of high- and low-ability workers in the population of interest to understand if this constraint could generate an overall increase in efficiency. On the other hand, imposing a deadline per task produces significant costs on the performance of both types of workers, although the reduction is less pronounced among the low-ability subjects.

Interestingly, this paper has causally identified an exogenous determinant of worker performance. Work schedules could indeed contextually increase or decrease the performance of different types of workers. This result has important implications, especially in situations where relative performance evaluations, such as promotion, retention, hiring or firing decisions, are used. In particular, the relative performance comparison among different candidates is partly determined by the interaction between the type of schedule each worker is typically assigned to, as well as workers' baseline skills.

Additionally, this study has important implications for the labor market matching problem. While standard models match workers to jobs, the results suggest that there are benefits from providing more autonomy to high-ability workers, even within the same job. Therefore,

there are significant efficiency gains in using the schedule of work as an additional matching dimension.

Taken together, this evidence is economically behavioral since it shows that certain types of individuals reach a higher optimum under a constrained maximization problem than under an unconstrained problem. From a theoretical perspective, these constraints could have been self-imposed. Thus, these findings support the idea that certain types of individuals may be less able to cope with this scheduling problem than we might have expected.

To conclude, this evidence suggests considering the work schedule as a relevant choice variable when considering the optimal job design. Moreover, since these results are in line with the evidence on task juggling (Coviello et al., 2015) and on multitasking (Buser and Peter, 2012), they further suggest that this type of inefficiency does not heavily depend on the type of sample or the type of task. Future studies expanding the research on this topic should account for individual heterogeneity in cognitive ability and further examine the persistence of these treatment effects in alternative settings, for example, with trained individuals.

References

- Adler, R. F. and Benbunan-Fich, R. (2012). Juggling on a high wire: Multitasking effects on performance. *International Journal of Human-Computer Studies*, 70(2):156–168.
- Aeon, B. and Aguinis, H. (2017). It’s about time: New perspectives and insights on time management. *The Academy of Management Perspectives*, pages amp–2016.
- Altman, D. G. and Royston, P. (2006). The cost of dichotomising continuous variables. *Bmj*, 332(7549):1080.
- Alubaydli, O. and List, J. A. (2015). On the generalizability of experimental results in economics. *Handbook of Experimental Economic Methodology*.
- Ariely, D. and Wertenbroch, K. (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological science*, 13(3):219–224.
- Bartling, B., Fehr, E., and Schmidt, K. M. (2012). Screening, competition, and job design: Economic origins of good jobs. *The American Economic Review*, 102(2):834–864.

- Bellemare, C., Bissonnette, L., and Kröger, S. (2016). Simulating power of economic experiments: the powerbbk package. *Journal of the Economic Science Association*, 2(2):157–168.
- Bisin, A. and Hyndman, K. (2014). Present-bias, procrastination and deadlines in a field experiment. Technical report, National Bureau of Economic Research.
- Brañas-Garza, P., Kujal, P., and Lenkei, B. (2019). Cognitive reflection test: Whom, how, when. *Journal of Behavioral and Experimental Economics*, 82:101455.
- Bray, R. L., Coviello, D., Ichino, A., and Persico, N. (2016). Multitasking, multiarmed bandits, and the italian judiciary. *Manufacturing & Service Operations Management*, 18(4):545–558.
- Brown, C. L., Kaur, S., Kingdon, G., and Schofield, H. (2022). Cognitive endurance as human capital. Technical report, National Bureau of Economic Research.
- Brynjolfsson, E., Horton, J. J., Ozimek, A., Rock, D., Sharma, G., and TuYe, H.-Y. (2020). Covid-19 and remote work: an early look at us data. *National Bureau of Economic Research*, w27344.
- Burgess, P. W. (2015). Serial versus concurrent multitasking: From lab to life. *The Handbook of Attention*, page 443.
- Buser, T. and Peter, N. (2012). Multitasking. *Experimental Economics*, 15(4):641–655.
- Camerer, C. F., Johnson, E., Rymon, T., and Sen, S. (1993). Cognition and framing in sequential bargaining for gains and losses. *Frontiers of game theory*, 1:27–47.
- Chandon, P., Hutchinson, J., Bradlow, E., and Young, S. H. (2006). Measuring the value of point-of-purchase marketing with commercial eye-tracking data.
- Chirkov, V. I., Ryan, R., and Sheldon, K. M. (2010). *Human autonomy in cross-cultural context: Perspectives on the psychology of agency, freedom, and well-being*, volume 1. Springer Science & Business Media.
- Claessens, B. J., Van Eerde, W., Rutte, C. G., and Roe, R. A. (2007). A review of the time management literature. *Personnel review*, 36(2):255–276.
- Corgnet, B., Gómez-Miñambres, J., and Hernán-Gonzalez, R. (2015a). Goal setting and monetary incentives: When large stakes are not enough. *Management Science*, 61(12):2926–2944.

- Corgnet, B., Gómez-Miñambres, J., and Hernán-Gonzalez, R. (2018). Goal setting in the principal-agent model: Weak incentives for strong performance. *Games and Economic Behavior*, 109:311–326.
- Corgnet, B., Gonzalez, R. H., and Mateo, R. (2015b). Cognitive reflection and the diligent worker: an experimental study of millennials. *PloS one*, 10(11):e0141243.
- Corgnet, B., Hernán-González, R., and Schniter, E. (2015c). Why real leisure really matters: Incentive effects on real effort in the laboratory. *Experimental Economics*, 18(2):284–301.
- Costa-Gomes, M., Crawford, V. P., and Broseta, B. (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica*, 69(5):1193–1235.
- Coviello, D., Ichino, A., and Persico, N. (2014). Time allocation and task juggling. *The American Economic Review*, 104(2):609–623.
- Coviello, D., Ichino, A., and Persico, N. (2015). The inefficiency of worker time use. *Journal of the European Economic Association*, 13(5):906–947.
- Davenport, T. H. (2005). *Thinking for a living: how to get better performances and results from knowledge workers*. Harvard Business Press.
- Dawling, P., Nagel, R., Camerer, C. F., and Rangel, A. (2011). Search dynamics in consumer choice under time pressure: An eye-tracking study. *The American Economic Review*, 101(2):900–926.
- De Paola, M. and Gioia, F. (2016). Who performs better under time pressure? results from a field experiment. *Journal of Economic Psychology*, 53:37–53.
- Duclos, R. (2015). The psychology of investment behavior:(de) biasing financial decision-making one graph at a time. *Journal of Consumer Psychology*, 25(2):317–325.
- Eerde, W. V. (2003). Procrastination at work and time management training. *The Journal of psychology*, 137(5):421–434.
- Falk, A. and Kosfeld, M. (2006). The hidden costs of control. *American Economic Review*, 96(5):1611–1630.

- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178.
- Foroughi, C. K., Werner, N. E., Nelson, E. T., and Boehm-Davis, D. A. (2014). Do interruptions affect quality of work? *Human factors*, 56(7):1262–1271.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, pages 25–42.
- Friedman, G. (2014). Workers without employers: shadow corporations and the rise of the gig economy. *Review of Keynesian Economics*, 2.2:171–188.
- Gabaix, X., Laibson, D., Moloche, G., and Weinberg, S. (2006). Costly information acquisition: Experimental analysis of a boundedly rational model. *The American Economic Review*, pages 1043–1068.
- González, V. M. and Mark, G. (2005). Managing currents of work: Multi-tasking among multiple collaborations. In *ECSCW 2005*, pages 143–162. Springer.
- Green, P. and Skinner, D. (2005). Does time management training work? an evaluation. *International Journal of Training and Development*, 9(2):124–139.
- Gregg, F. and Crawford, V. P. (2006). Cognition and behavior in two-person guessing games: An experimental study. *The American economic review*, 96(5):1737–1768.
- Häfner, A. and Stock, A. (2010). Time management training and perceived control of time at work. *The journal of psychology*, 144(5):429–447.
- Howard, A. E. (1995). *The changing nature of work*. Jossey-Bass.
- Hüsser, A. and Wirth, W. (2014). Do investors show an attentional bias toward past performance? an eye-tracking experiment on visual attention to mutual fund disclosures in simplified fund prospectuses. *Journal of Financial Services Marketing*, 19(3):169–185.
- Johnson, E. J., Camerer, C., Sen, S., and Rymon, T. (2002). Detecting failures of backward induction: Monitoring information search in sequential bargaining. *Journal of Economic Theory*, 104(1):16–47.

- Kocher, M. G. and Sutter, M. (2006). Time is moneytime pressure, incentives, and the quality of decision-making. *Journal of Economic Behavior & Organization*, 61(3):375–392.
- Levitt, S. D. and List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *The journal of economic perspectives*, 21(2):153–174.
- Lyness, K. S., Gornick, J. C., Stone, P., and Grotto, A. R. (2012). Its all about control: Worker control over schedule and hours in cross-national context. *American Sociological Review*, 77(6):1023–1049.
- Macan, T. H. (1994). Time management: Test of a process model. *Journal of applied psychology*, 79(3):381.
- Macan, T. H. (1996). Time-management training: Effects on time behaviors, attitudes, and job performance. *The Journal of psychology*, 130(3):229–236.
- Monahan, K. (2018). *How Behavioral Economics Influences Management Decision-making: A New Paradigm*. Academic Press.
- Monsell, S. (2003). Task switching. *Trends in cognitive sciences*, 7(3):134–140.
- Monsell, S. and Driver, J. (2000). Banishing the control homunculus. *Control of cognitive processes: Attention and performance XVIII*, pages 3–32.
- Morgan, J. (2014). *The future of work: Attract new talent, build better leaders, and create a competitive organization*. John Wiley & Sons.
- Olson, M. H. (1983). Remote office work: changing work patterns in space and time. *Communications of the ACM*, 26(3):182–187.
- Oud, B., Krajbich, I., Miller, K., Cheong, J., Botvinick, M., and Fehr, E. (2016). Irrational time allocation in decision-making. *Proceedings. Biological sciences*, 283(1822).
- Payne, J. W., Bettman, J. R., and Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge University Press.

- Peterson, N. G., Mumford, M. D., Borman, W. C., Jeanneret, P., and Fleishman, E. A. (1999). *An occupational information system for the 21st century: The development of O* NET*. American Psychological Association.
- Pinedo Caro, L., O'Higgins, N., and Berg, J. (2021). Young people and the gig economy. *International Labour Organization*.
- Pink, D. H. (2011). *Drive: The surprising truth about what motivates us*. Penguin.
- Rogers, R. D. and Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of experimental psychology: General*, 124(2):207.
- Toplak, M. E., West, R. F., and Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Memory & cognition*, 39(7):1275.
- Tucker, A. L. (2007). An empirical study of system improvement by frontline employees in hospital units. *Manufacturing & Service Operations Management*, 9(4):492–505.
- Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1):75–98.

6 Appendix A

Table A1: Summary Statistics

Variable	N	Mean	St. deviation	Min	Max
P(Complete a Task)	2958	0.63	0.48	0	1
Score	87	14.52	6.21	3	29
Gender	87	0.25	0.44	0	1
Eco. Courses	87	0.07	0.25	0	1
Risk	87	6.22	1.51	3	9
CRT_1	87	0.24	0.43	0	1
CRT_2	87	0.28	0.45	0	1
CRT_3	87	0.25	0.44	0	1

Notes: *P(Complete one Task)* measures the average probability of completing one task. *Score* sums all points obtained in the two parts of the test. *Gender* equals to one for male. *Eco. Courses* indicates the fraction of subjects who have taken part in any economic course at the time of the experiment. *Risk* is a measure of individual risk attitudes created using the answer to the following question: “In general, are you a person ready to take risks, or do you avoid to take risks? Please indicate your answer on a scale from 1 to 10, where 1 means that you do not want to take risk and 10 means that you are ready to take risks”. The three *CRT* measures derive from the answers to the three questions of the Cognitive Reflection Test in [Frederick \(2005\)](#).

Table A2: Test on Random Allocation

Variable	Baseline Mean	Fixed Sequence Mean	Fixed Time Mean	p-value(*) columns 1 & 2	p-value(*) columns 2 & 3
P(Complete a Task)	0.662	0.686	0.631	0.530	0.688
Score (First Part)	10	9.75	9.70	0.7077	0.9879
Gender	0.03	0.10	0.06	0.128	0.217
Eco. Courses	0.03	0.10	0.06	0.2718	0.5369
Risk	6.13	6.31	6.24	0.2824	0.7097
CRT_1	0.33	0.23	0.21	0.4602	0.4156
CRT_2	0.33	0.29	0.21	0.6978	0.2789
CRT_3	0.33	0.23	0.25	0.4602	0.6156

Notes: (*) Two-sample Wilcoxon Mann-Whitney test. *Score (First Part)* measures the performance obtained in the first part of the experiment.

Table A3: Scheduling Measures by Skill Type

	Mean	St. deviation	Min	Max
<i>High-skill</i>				
P(Complete one Task)	0.69	0.46	0	1
P(Mistake in the Task)	0.25	0.43	0	1
Lookups per question	2.86	1.95	0	15
Average time on task	115.93	111.23	14.72	898.81
Time for wrong	171.86	136.92	27.61	898.81
Time for right	97.53	94.55	14.72	725.86
Time for missing	176.91	148.77	16.75	536.56
<i>Low-skill</i>				
P(Complete one Task)	0.60	0.49	0	1
P(Mistake in the Task)	0.27	0.44	0	1
Lookups per question	2.84	1.92	1	17
Average time on task	121.02	103.97	9.75	653.34
Time for wrong	155.48	109.22	25.92	606.41
Time for right	108.11	98.98	9.75	653.34
Time for missing	174.43	145.29	9.75	653.343

Notes: All the above information are related to subject behavior in the first part of the experiment. *P(Complete one Task)* measures the average probability of completing one task. *P(Mistake in the Task)* measure the probability of making a mistake in the task. *Lookups per question* is the number of times the subject has looked at the same task before giving the final answer. *Average time on task* measures the mean time spent on each task. *Time for wrong* indicates the average time spent on tasks that were not correctly completed. *Time for right* indicates the average time spent on correctly completed tasks. *Time for missing* indicates the average time spent on non-completed tasks. Notice that all the time measures are expressed in seconds.

Table A4: Cognitive Types Distribution

Number of Wrong Answers in the CRT (categorical)	Freq.	Percent	Cum.
3	44	50.57	50.57
2	23	26.44	77.01
1	16	18.39	95.4
0	4	4.6	100
Cognitive Types (Indicator)			
1 (Low Skilled)	44	50.57	50.57
0 (High Skilled)	43	49.43	100
Total	87	100	

Notes: The categorical variable shows the number of wrong answers in the 3-items CRT. A value of three indicates that the subject has provided a wrong answer in all the questions of the CRT, while zero indicates that the subject always provided a correct answer. Note that the indicator variable is equal to 1 for the subjects who wrongly answered all the questions of the CRT (Low Skilled Types).

Table A5: Behavior of High/Low Skilled Types in the Unconstrained group across parts

	High Skill			Low Skill		
	Lookups	Time on Task	P(Mistake)	Lookups	Time on Task	P(Mistake)
Second Part	-0.052 (0.177)	2.235 (2.592)	-0.121*** (0.034)	-0.054 (0.260)	-4.782 (3.136)	0.118** (0.045)
Constant	3.993*** (0.089)	130.886*** (1.296)	0.288*** (0.017)	4.255*** (0.130)	142.277*** (1.568)	0.260*** (0.023)
Observations	612	612	612	408	408	408
Individual fixed effects	✓	✓	✓	✓	✓	✓
Section fixed effect	✓	✓	✓	✓	✓	✓
Clustered Std. errors	✓	✓	✓	✓	✓	✓
<i>Adj.R</i> ₂	0.28	0.010	0.10	0.24	-0.027	0.031

Notes: *Lookups*, *Time on Task*, *P(Mistake)* are the outcome variables. *Lookups* counts the number of times the task was seen by the same subject before completion, over time. *Time on Task* measures the seconds spent on each task, over time. *P(Mistake)* indicates the probability of making a mistake in the task, over time. *Second Part* is a dummy variable equal to one for the second part of the experiment. The regressions include individual, and section fixed effects. Errors are clustered at the individual level. Standard errors reported in parentheses, and *, **, and *** indicate statistical significance at the 10%, the 5%, and the 1% level, respectively.

3D Plots on the Number of Lookups per Task over Time by Types

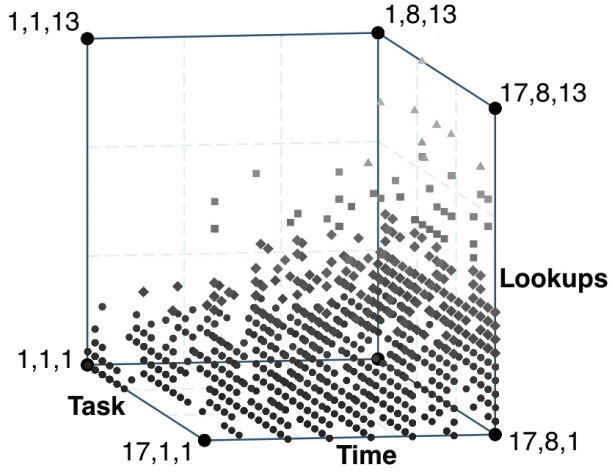


Figure A1: High Ability - Lookups

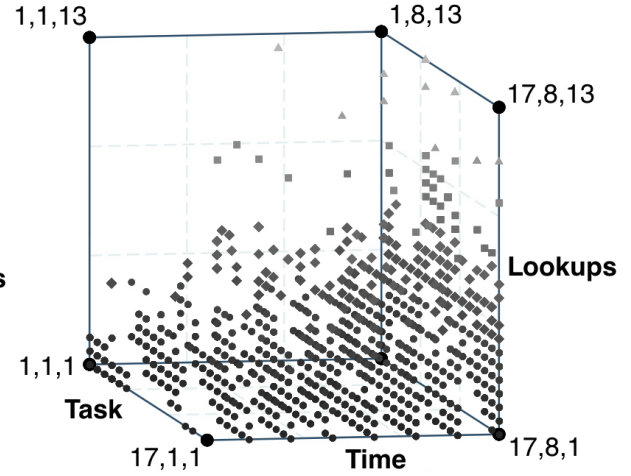


Figure A2: Low Ability - Lookups

3D Plots on the Time spent on Tasks over Time by Types

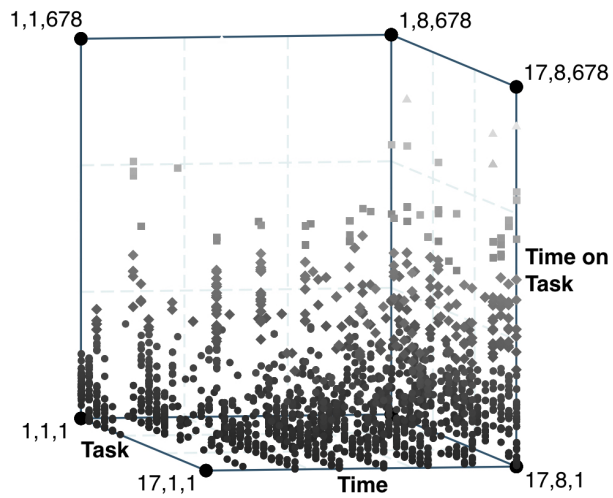


Figure A3: High Ability - Time Allocation

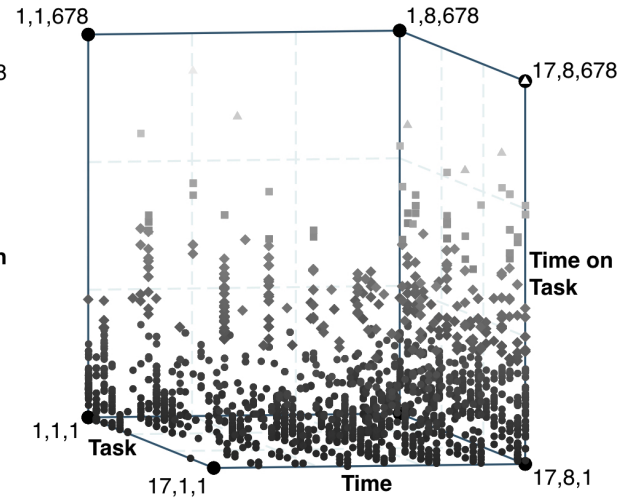


Figure A4: Low Ability - Time Allocation

Table A6: Performance in the First Part by High/Low Skill Types

	(1)	(2)	(3)	(4)	(5)
<i>Low Skill</i>	-0.077*** (0.024)	-0.139*** (0.050)	-0.137*** (0.049)	-0.072 (0.098)	
t_2		0.089 (0.077)	0.097 (0.075)	-0.087 (0.138)	0.056 (0.081)
t_2		0.058 (0.060)	0.055 (0.058)	-0.024 (0.126)	0.079 (0.071)
<i>High Skill</i>					
t_2		-0.071 (0.049)	-0.050 (0.045)	0.016 (0.092)	-0.006 (0.052)
t_3		-0.224*** (0.040)	-0.147*** (0.036)	-0.151** (0.072)	-0.193*** (0.051)
Constant	0.328 (0.296)	0.655** (0.269)	0.540** (0.250)	0.657** (0.261)	0.721*** (0.031)
Observations	1479	1479	1479	1479	1479
Controls	✓	✓	✓	✓	✓
Time fixed effect		✓	✓	✓	✓
Section fixed effects			✓		✓
Behavioral Controls				✓	
Individual fixed effects					✓
$Adj.R_2$	0.010	0.042	0.111	0.130	0.114

Notes: $P(\text{Complete one Task})$ is the outcome variable in each column. *Low Skill* is a dummy variable equal to one if the subject has wrongly answered all the 3 questions of the CRT. *Controls* include observable characteristics: gender, age, risk attitudes, background in economics, instructions comprehension. *Behavioral Controls* control for the number of lookups and the time spent on each task. The regression results are estimated using the model of equation 1. The coefficients $t_2 - t_3$ in the Low Skill panel shows the interactions of time fixed effects with the Low-skill dummy. Notice that the time fixed effects $t_1 - t_3$ are constructed so that each time effect represents a range of time of 15 minutes. Errors are clustered at the individual and session levels. Standard errors reported in parentheses, and *, **, and *** indicate statistical significance at the 10%, the 5%, and the 1% levels, respectively

Table A7: Correlation between Performance and Behavior

	Full		Unconstrained		Fixed Sequence		Fixed Time	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Second Part	0.023 (0.041)	0.027 (0.058)	0.116* (0.034)	0.104 (0.045)	-0.030 (0.020)	-0.036 (0.054)	-0.119* (0.031)	-0.020 (0.082)
Low Skill	-0.081** (0.025)	-0.094 (0.060)	-0.001 (0.007)	0.008 (0.025)	-0.176** (0.040)	-0.031 (0.046)	-0.141* (0.037)	-0.030 (0.093)
Low Skill x Second Part	-0.117** (0.048)	-0.090 (0.080)	-0.224*** (0.018)	-0.231** (0.044)	-0.004 (0.075)	-0.148 (0.073)	-0.021 (0.019)	-0.125 (0.082)
Time Correlation		0.131 (0.216)		0.036 (0.551)		0.231 (0.124)		0.724 (0.468)
Second Part x Time Correlation		-0.627* (0.306)		-0.291 (0.159)		-0.336 (0.229)		-0.862 (0.645)
Mean Lookups		-0.010 (0.018)		0.010 (0.035)		-0.003 (0.031)		-0.102 (0.043)
Second Part x Mean Lookups		-0.007 (0.022)		-0.091 (0.045)		-0.201** (0.034)		0.115* (0.035)
Low Skill x Time Correlation		-0.168 (0.329)		0.459 (0.978)		-0.318 (0.442)		-1.057 (0.513)
Low Skill x Mean Lookups		0.008 (0.034)		-0.017 (0.066)		-0.087** (0.019)		0.142 (0.089)
Low Skill x Second Part x Time Correlation		1.496* (0.658)		1.024 (0.685)		0.560 (0.504)		2.174 (1.817)
Low Skill x Second Part x Mean Lookups		-0.017 (0.036)		0.024 (0.062)		0.407** (0.062)		-0.148* (0.048)
Constant	0.658 (0.367)	0.701* (0.337)	0.724 (0.482)	0.841 (0.395)	2.724** (0.539)	2.697 (1.098)	-0.532 (0.870)	-0.952 (1.220)
Observations	2958	2958	1020	1020	952	952	986	986
Controls	✓	✓	✓	✓	✓	✓	✓	✓
Section fixed effect	✓	✓	✓	✓	✓	✓	✓	✓
Clustered Std. errors	✓	✓	✓	✓	✓	✓	✓	✓
Behavioral Controls		✓		✓		✓		✓
Adj. R ₂	0.12	0.12	0.12	0.12	0.12	0.12	0.13	0.13

Notes: *P(Complete one Task)* is the outcome variable. *Low Skill* is a dummy variable equals to one if the subject has wrongly answered all the 3 questions of the CRT. *Mean Lookups* counts the average number of times the subject has switched task (before completion) in each part. *Time Correlation* measures the correlation between the time spent on each task and the time passed by. *Second Part* is a dummy variable equal to one if the second part of the experiment is considered. Errors are clustered at the individual and session levels. Standard errors reported in parentheses, and *, **, and *** indicate statistical significance at the 10%, the 5%, and the 1% level, respectively

Table A8: Heterogeneous Treatment Effects by Types - Other Outcomes

	Final Score	P(Mistake in the Task)	P(Not Completing one Task)
<i>Low Skill</i>			
Low Skill x Second Part x Unconstrained	-4.708*** (0.444)	0.243*** (0.054)	0.005 (0.010)
Low Skill x Second Part x Fixed Sequence	4.661*** (1.384)	-0.241*** (0.076)	0.026 (0.023)
Low Skill x Second Part x Fixed Time	4.069*** (0.576)	-0.173** (0.072)	-0.016 (0.025)
<i>High Skill</i>			
Second Part x Unconstrained	2.458*** (0.628)	-0.123*** (0.033)	0.005 (0.004)
Second Part x Fixed Sequence	-1.974** (0.660)	0.107** (0.045)	0.000 (0.006)
Second Part x Fixed Time	-4.569*** (0.831)	0.151*** (0.045)	0.028 (0.020)
Constant	9.848*** (0.137)	0.249*** (0.020)	0.009* (0.005)
Observations	174	2958	2958
Time fixed effects		✓	✓
Individual fixed effects		✓	✓
Section fixed effects		✓	✓
Adj.R ₂	0.570	0.056	0.118

Notes: *Final Score*, *P(Mistake in the Task)* and *P(Not Completing one Task)* are the outcome variables, indicating the sum of points obtained from the completed tasks, the probability of having mistakenly completed, and of not having completed a task, respectively. *Low Skill* is an indicator variable equal to one if the subject has wrongly answered all the 3 questions of the CRT. *Fixed Time* and *Fixed Sequence* are the dummy variables indicating the treatment groups. *Second Part* is a dummy variable, which equals one if the second part of the experiment is considered. Errors are clustered at the individual and session levels. Standard errors reported in parentheses, and *, **, and *** indicate statistical significance at the 10%, the 5%, and the 1% levels, respectively

Table A9: Treatment Effects by Continuous Behavioral Measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Second Part x Unconstrained	0.318*** (0.116)	0.014 (0.108)	0.319** (0.154)	0.321*** (0.116)	0.017 (0.111)	0.327** (0.152)	0.321** (0.128)	0.017 (0.111)	0.327* (0.166)
Second Part x Fixed Sequence	-0.362** (0.173)	-0.064 (0.124)	-0.373* (0.207)	-0.360** (0.171)	-0.067 (0.129)	-0.377* (0.204)	-0.360** (0.168)	-0.067 (0.129)	-0.377* (0.200)
Second Part x Fixed Time	-0.508*** (0.150)	-0.159 (0.130)	-0.503*** (0.192)	-0.506*** (0.148)	-0.160 (0.132)	-0.505*** (0.189)	-0.506*** (0.153)	-0.160 (0.132)	-0.505** (0.200)
Second Part x Fixed Sequence x Lookups	0.078** (0.040)		0.072 (0.048)	0.076* (0.039)		0.069 (0.047)	0.076** (0.037)		0.069 (0.042)
Second Part x Fixed Time x Lookups	0.083** (0.035)		0.084** (0.037)	0.082** (0.035)		0.083** (0.037)	0.082** (0.035)		0.083** (0.035)
Second Part x Fixed Sequence x Time Allocation		0.000 (0.001)	0.000 (0.001)		0.000 (0.001)	0.000 (0.001)		0.000 (0.001)	0.000 (0.001)
Second Part x Fixed Time x Time Allocation		-0.000 (0.001)	-0.000 (0.001)		-0.000 (0.001)	-0.000 (0.001)		-0.000 (0.001)	-0.000 (0.001)
Observations	2958	2958	2958	2958	2958	2958	2958	2958	2958
Section fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Time fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Individual fixed effects				✓	✓	✓	✓	✓	✓
Clustered Errors							✓	✓	✓
Adj.R ₂	0.114	0.112	0.117	0.144	0.142	0.143	0.144	0.142	0.143

Notes: Estimation of treatment effects using continuous behavioral measures as a way to identify the worker type. *P(Completing one Task)* is the outcome variable in each column. *Lookups* is a continuous variable counting the mean number of times the subject has looked the incorrect tasks. *Time Allocation* is a continuous variable measuring the mean time spent on incorrect tasks. *Fixed Time* and *Fixed Sequence* are the dummy variables indicating the treatment groups. *Second Part* is a dummy variable, which equals one if the second part of the experiment is considered. Errors are clustered at the individual and session levels. Standard errors reported in parentheses, and *, **, and *** indicate statistical significance at the 10%, the 5%, and the 1% levels, respectively.

Table A10: Heterogeneous Treatment Effects by Types - Logistic Panel Regressions

	P(Complete one Task)	P(Complete one Task)	P(Complete one Task)
<i>High Skill</i>			
Second Part x Unconstrained	0.658*** (0.196)	0.683*** (0.197)	0.697*** (0.199)
Second Part x Fixed Sequence	-0.543* (0.277)	-0.696** (0.281)	-0.722** (0.282)
Second Part x Fixed Time	-1.269*** (0.328)	-1.333*** (0.330)	-1.343*** (0.329)
<i>Low Skill</i>			
Second Part x Unconstrained	-1.134*** (0.290)	-1.185*** (0.292)	-1.188*** (0.291)
Second Part x Fixed Sequence	1.110*** (0.410)	1.144*** (0.413)	1.144*** (0.413)
Second Part x Fixed Time	1.115*** (0.425)	1.198*** (0.427)	1.192*** (0.426)
Observations	2958	2958	2958
Section fixed effects	✓	✓	✓
Time fixed effects		✓	✓
Individual fixed effects			✓

Notes: Logistic panel regressions. *P(Complete one Task)* is the outcome variable. *Low Skill* is an indicator variable equal to one if the subject has wrongly answered all the 3 questions of the CRT. *Fixed Time* and *Fixed Sequence* are the dummy variables indicating the treatment groups. *Second Part* is a dummy variable, which equals one if the second part of the experiment is considered. Standard errors reported in parentheses, and *, **, and *** indicate statistical significance at the 10%, the 5%, and the 1% level, respectively

Table A11: Marginal Effects of Logistic Panel Regressions

<i>High Skill</i>	
Second Part x Unconstrained	0.147*** (0.041)
Second Part x Fixed Sequence	-0.076*** (0.029)
Second Part x Fixed Time	-0.094*** (0.023)
<i>Low Skill</i>	
Second Part x Unconstrained	-0.250*** (0.060)
Second Part x Fixed Sequence	0.120*** (0.043)
Second Part x Fixed Time	0.083*** (0.029)

Notes: Marginal effects. $P(\text{Complete one Task})$ is the outcome variable. *Low Skill* is an indicator variable equal to one if the subject has wrongly answered all the 3 questions of the CRT. *Fixed Time* and *Fixed Sequence* are the dummy variables indicating the treatment groups. *Second Part* is a dummy variable, which equals one if the second part of the experiment is considered. Standard errors reported in parentheses, and *, **, and *** indicate statistical significance at the 10%, the 5%, and the 1% levels, respectively

7 Data Availability Statement

The anonymous data that support the findings of this study are openly available.

8 Appendix B

8.1 Instructions

Welcome!

Thanks for your participation. This study aims to understand how people make decisions. The study was funded by the University of Bologna.

As a show-up fee you earn 8€. During the session, you can earn more money depending on your choices. Your earnings will be expressed in points, and they will be converted into euros at the rate of 2 points = 1 euro. Payment will be made at the end of the session and privately.

At the end of the session, please remain seated. The researcher will come to your workstation to hand you a private envelope containing your payment.

During the study, we do not allow any communication with other participants. You should turn off your phone now. For any question, please raise your hand and we will come to your workstation.

In this study, you will face a training test, whose structure and evaluation resemble the ones of the admission test to the Faculty of Economics and Statistics.

Now we will start to describe in detail the structure of the test and how you can use the computer in front of you to answer it.

How the test is structured:

The test is divided into two parts: Part 1 and Part 2.

Each part is composed of three thematic areas: Mathematics, Logic, and Verbal Comprehension.

The table below illustrates in detail the content of the test.

Each part includes 17 questions, divided into the three sections as shown in Table 1. In particular:

Table 1: Structure of the Test

Test Part 1				
Section	Number of tasks	Time available	Minimum Score	Maximum Score
Logic	6		-1.5	6
Verbal Comprehension	5		-1.25	5
Math	6		-1.5	6
Total	17	45	-4.25	17

Test Part 2				
Section	Number of tasks	Time available	Minimum Score	Maximum Score
Logic	6		-1.5	6
Verbal Comprehension	5		-1.25	5
Math	6		-1.5	6
Total	17	45	-4.25	17

- The first section contains 6 questions of logic.
- The second section includes 5 questions of verbal comprehension. This section requires reading a text on a general topic. The text is followed by a series of questions, whose answers must be inferred exclusively from the content of the text.
- The third section contains 6 questions of math.

The questions are structured as multiple choices tasks. Each question has 5 possible answers, of which only one provides the correct solution.

For each question, it is assigned 1 point for the correct answer, 0 points for the missing and -0.25 points for the wrong.

Table 1 shows the minimum and maximum possible score for each Part and each section.

We will move now to a practical example that will show you how the test will appear on your computer and how you can select / de-select questions and answers.

Practical example:

As you can see, the first screen shows the questions of the logic section.

In general, to understand which section or which question is “active” you just have to see which of the sections’ box is coloured and what number is enlightened among the array of buttons on the bottom left of the screen.

In the centre of the screen, you will see the number of the task and the text of the question. There are 5 multiple choices for answering each task, and each option is associated with a round selector.

In this example, we assume that you think that the right answer for question 1 is the second option. To select it, you have to position within the associated selector and click the left button of the mouse. As you can see once you have clicked on it, the round selector and the button of the task became coloured. This fact indicates which question you answered and which answer you have chosen. To change your answer, just click on the selector associated with the new option you want to select. Try to choose answer 5. To clear the answer, just click back into the selector of the previously selected answer. Try to clear answer 5. You can now choose a new answer or choose to do not answer to this question. Suppose you choose to omit the answer for question 1, then go to question 3. As you can see, the box of task 1 is uncoloured, precisely because you have chosen to omit the answer.

You can choose to go back and forth to check the answers to all the questions of the various sections.

Let us now turn to the Verbal Comprehension section. Click on the button of the section. As you can see, the section box is coloured now. You can see the text to read and a “T” in the bottom left corner, indicating that you are on the text screen.

Also in this section, as well as in Mathematics, you can answer to questions as explained in the previous example for the Logic section.

Now, look at the time bar. In the right part of the screen, there is a time scrolling bar which will be always displayed during the test and which tells you the minutes and seconds remaining until the end of the test. The bar has a mark every ten seconds: 60 seconds, 50 seconds, 40

seconds and so on. This bar is to show approximately how many seconds you have left.
Write now, on the instruction sheet, the remaining time at this moment.

Time remaining: minutes and seconds

Please note:

At the last minute, it will appear that the remaining minutes are zero but notice that you will still have 60 seconds to use, which will scroll down as time passes.

In Figure 2, you can see that about 15 seconds are left.

At any time you can click on the “End” button which will stop the test and will bring you to the final results page. The test is also stopped automatically at the end of the available time.

Press now the “End” button.

We finished the practical example.

This study consists of two parts. Now we read the instructions relating to the first part. Please, pay attention.

Figure 2



Part 1

In this first part, you have 45 minutes to answer 17 questions.

The questions are organized following the structure presented above. During the test, you can move back and forth, both within and across sections and you can freely decide which questions you want to answer first and how much time you want to spend on each task, with a total available time equals to 45 minutes.

The final score is the sum of points earned by correctly answer the task minus the penalty for wrong answers, as shown above.

If there are any questions, please raise your hand, and we will come to your workstation.

Now we will start the first part of the study.

Part 2 - Fixed Time

The second part of the study is identical to the first one with just one difference.

In the first part, you were free to decide in which sequence you want to answer the questions and how much time you want to spend on each of them.

In this part of the study, you will only be able to freely decide in which order you want to answer the questions, but you will have a maximum time for each question of 2 minutes and 22 seconds. In particular, you can still choose to switch freely among questions, even among the various sections, but you can not choose how much time you want to allocate to each question.

In particular, if you change question before the time for that task is finished, the timer will stop exactly when you switched the question; in this way, you will not lose the remaining seconds and you can re-use them as soon as you return to the same question.

When the time for one question ends, you will see a message on the screen asking to choose another task.

The score is always computed in the same way: 1 point for each correct answer, -0.25 for each wrong answer, 0 for missing.

If there are any questions, please raise your hand, and we will come to your workstation.

Now we will start the second part of the study.

Part 2 - Fixed Sequence

The second part of the study, is identical to the first one with just one difference.

In the first part, you were free to decide in which sequence you want to answer the questions and how much time you want to spend on each of them.

In this part of the study, you will be able to decide freely just how you want to allocate the time across questions, while the sequence of questions is fixed. In particular, you can not choose to switch freely back and forth from one question to another, but you have to answer the test following a given order of the tasks: from task 1 to task 2 to task 3, starting from the Logic section, then moving to the Verbal and finally to the Math. Once you choose a section, you have to answer the tasks of that section in order of ascending numbers.

The score is always computed in the same way: 1 point for each correct answer, -0.25 for each wrong answer, 0 for missing.

If there are any questions, please raise your hand, and we will come to your workstation.

Now we will start the second part of the study.

8.2 Questionnaire

We now kindly ask you to complete this questionnaire.

The answers in this section will not affect your final score.

Some of these questions are related to personal information that will be useful for this study.

Your identity will not be revealed under any circumstances.

Please answer carefully.

Once you answered, you can not edit the answer.

Press OK to start.

Thank you!

- 1 Sex (Press the related button)
- 2 Age (Move your red triangle and press OK to confirm)
- 3 Have you ever attended economics courses?
- 4 Have you ever taken part in other researches within the university? (Select one of the answers and click OK)
- 5 In general, are you a person ready to take risks, or not? Please indicate your answer on a scale of 1 to 10, where 1 means "I always prefer to do not take risks" and 10 means that risks "I am always ready to take risks."
- 6 In which part do you think to have reached higher performance? Please select your answer.

8.3 Cognitive Reflection Test:

- 1 A bat and a ball cost \$ 1.10 in total. The bat costs \$ 1.00 blackberries than the ball. How much does the ball cost?
- 2 If it takes 5 machines 5 minutes to make five widgets, how long would it take 100 machines to make 100 widgets?

- 3 In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?