



**UNIVERSITY
OF TURKU**

Pressure in the Spotlight: Effects of Monitoring Pressure and Outcome Pressure on Time-Sharing Performance

Psychology

Master's thesis

Department of Psychology and Speech-Language Pathology

Author:

Niki Pennanen

28.10.2024

Turku

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin Originality Check service.

Master's thesis

Subject: Psychology

Author(s): Niki Pennanen

Title: Pressure in the Spotlight: Effects of Monitoring Pressure and Outcome Pressure on Time-Sharing Performance

Supervisor(s): Docent Lauri Oksama

Number of pages: 36 pages

Date: 28.10.2024

Performing under pressure, particularly in multitasking environments, is a critical challenge in both everyday life and high-stakes professions. This study investigated the differential effects of monitoring and outcome pressure on time-sharing performance and attentional allocation. Using a within-subjects design, 30 participants completed a cognitive task under three different pressure conditions while we recorded their eye movements. We hypothesized that in a high-demand time-sharing environment, outcome pressure would impair task performance and prioritization more significantly than monitoring pressure. However, our confirmatory analyses found no evidence to support this hypothesis. Interestingly, our additional exploratory analyses revealed that monitoring pressure, rather than outcome pressure, led to a statistically significant decrease in task performance. This unexpected finding is likely due to the sensorimotor demands of the task, specifically the need for precise and rapid mouse movements, which may have been disrupted by the participants' heightened self-consciousness under monitoring pressure.

Our findings contribute to the growing body of literature on the differential effects of monitoring and outcome pressure. Moreover, this study highlights the importance of carefully considering the influence of experimenter presence and other monitoring elements, as they potentially may inadvertently affect performance in computer-based tasks, even in studies not explicitly focused on psychological pressure. These insights underscore the complexity of psychological pressure effects and provide a foundation for refining experimental designs and exploring the distinct roles of different types of pressure in future research and practical applications.

Key words: Attentional control, time-sharing, multitasking, performance pressure, monitoring pressure, outcome pressure, eye movements.

Table of contents

1	Introduction	4
1.1	Attentional control and time-sharing	4
1.2	Mechanisms of pressure	5
1.2.1	Monitoring pressure	6
1.2.2	Outcome pressure	6
1.2.3	Previous studies comparing pressures	7
1.3	Goals of the study	8
1.3.1	Hypotheses	9
2	Methods	10
2.1	Participants	10
2.2	Stimuli	10
2.3	Measures	12
2.4	Design and Procedure	13
2.4.1	No-pressure condition	14
2.4.2	Outcome pressure condition	14
2.4.3	Monitoring pressure condition	14
3	Results	16
3.1	Pressure manipulation	16
3.2	Confirmatory Analysis	17
3.3	Exploratory Analysis	20
4	Discussion	24
4.1	Study Limitations	28
4.2	Conclusions	29
	References	31

1 Introduction

In today's fast-paced and demanding world, the ability to perform multiple tasks simultaneously in rapid and dynamic environments has become increasingly important. Consider the demands faced by air traffic controllers, surgeons in operating rooms, drivers at busy intersections, or chefs in restaurants who juggle multiple orders while maintaining quality. These scenarios demand quick and accurate allocation of attention across various stimuli, requiring exceptional prioritization and multitasking capabilities.

Performing under pressure is not just a concern for athletes or other high-stakes professions, it is an integral part of everyday life. Psychological pressure, stemming from time constraints, high expectations, being evaluated, or complex environments, can significantly impact cognitive processes such as attentional control and time-sharing. This pressure can lead to the phenomenon known as "choking", where individual performs significantly below their true skill level (Beilock & Gray, 2007). Attentional control, the ability to focus cognitive resources on relevant stimuli while ignoring distractions, is critical for efficient task performance (Eysenck et al., 2007). It is also crucial for time-sharing, the ability to allocate time and attention between concurrent tasks, which is a fundamental component of multitasking (C. D. Wickens, 1991). Understanding how psychological pressure influences these cognitive processes is important not only for theoretical insights but also for practical applications aimed at improving performance in both high-pressure and routine settings.

This study investigates the effects of different categories of psychological pressure on time-sharing performance and the allocation of attention. We employ a demanding time-sharing task and expose participants to two types of pressure, monitoring pressure and outcome pressure, separately. Our focus is to determine through task performance and eye movements whether different types of pressure situations produce distinct effects on performance and attentional processes.

1.1 Attentional control and time-sharing

Attentional control has been referred to by various terms (Burgoyne & Engle, 2020), including executive control (Baddeley, 1996), cognitive control (Botvinick et al., 2001), and executive attention (Engle, 2002). According to Wickens (2021), attentional control can be understood through two main concepts: the filter of selective attention and the allocation of cognitive resources. The filter allows individuals to concentrate on specific tasks by managing

the influx of information, while the resource allocation perspective considers how attentional capacity is distributed among concurrent tasks, especially when task demands are high. Regardless of the terminology or exact underlying definition, attentional control is considered crucial for many complex cognitive tasks.

The related concept of time-sharing is defined as “the process of rapidly switching attention from one task to another when two or more tasks are performed concurrently” (*APA Dictionary of Psychology*, n.d.). Time-sharing is often considered a process that facilitates multitasking, and the terms are frequently used interchangeably. When attempting to perform multiple overlapping tasks simultaneously, the primary challenge is resolving resource conflicts and properly prioritizing tasks. The exact mechanisms by which time-sharing processes resolve these resource problems is debated, with two main views presented. One view suggests that a strong executive component of attention directs focus to the most important tasks (e.g., Baddeley, 1996; Meyer & Kieras, 1997; Norman & Shallice, 1986). The other view, labelled “threaded cognition”, hypothesizes that a specialized executive process is not needed and that different tasks freely negotiate and resolve resource conflicts locally (Salvucci & Taatgen, 2008). For a detailed comparison and empirical study of these views, see Kulomäki et al. (2022).

In this study, we use time-sharing as a platform to investigate how different types of psychological pressure affect performance and allocation of attention during multitasking.

1.2 Mechanisms of pressure

Researchers have long acknowledged the associations between attentional processes and anxiety (e.g., Beck & Clark, 1997; Easterbrook, 1959; Eysenck et al., 2007), often triggered by pressure. High-stakes situations are known to cause performance pressure, which sometimes leads to worse-than-expected outcomes. Traditionally, there has been two different schools of thought to explain why pressure sometimes leads to poor performance in both cognitive and motor tasks (DeCaro et al., 2011). Distraction theories suggest that performance deteriorates under pressure due to task-irrelevant thoughts and worries capturing attention away from the task (Beilock & Carr, 2001; Lewis & Linder, 1997; Wine, 1971). In contrast, explicit monitoring theories argue essentially the opposite: pressure shifts too much attention towards the skill processes and procedures, disrupting their execution (Baumeister, 1984; Beilock & Carr, 2001; R. S. W. Masters, 1992). To address how pressure could both divert attention away from and towards the task at hand, DeCaro et al. (2011) focused on the

elements of the pressure situations themselves. While many real-world environments may contain mixed pressure elements, they identified two main categories of performance pressure potentially harming performance: monitoring pressure and outcome pressure.

1.2.1 Monitoring pressure

Monitoring pressure arises from performing while being observed and potentially evaluated by others, such as a teacher, audience, or video cameras (DeCaro et al., 2011). This feeling of being watched and evaluated shifts the individual's focus of attention more towards the skill processes and step-to-step procedures being performed. This can then lead to poorer performance if these processes are typically executed almost automatically, outside of awareness.

This route to failure has strong theoretical foundations in explicit monitoring, or self-focus, theories (e.g., Baumeister, 1984; Beilock & Carr, 2001; R. Masters & Maxwell, 2008; R. S. W. Masters, 1992). Most research on this type of skill failure under pressure has been conducted with sensorimotor skills, but support has also been found in cognitive tasks without motor components (DeCaro et al., 2011). A common feature of these tasks is that they involve highly proceduralized processes requiring little to no attentional control, as they are normally automated to a degree (e.g., Beilock & Carr, 2001; Jackson et al., 2006; R. S. W. Masters, 1992). Overall, monitoring pressure seems to cause "paralysis by analysis", where constant attempts to control skill execution disrupt normally fluent processes. Therefore, we expect that monitoring pressure will not significantly affect performance in our experimental task, which is heavily reliant on conscious attentional control instead of proceduralized processes.

1.2.2 Outcome pressure

Outcome pressure can be induced when the individual is offered an incentive to achieve a specific outcome (DeCaro et al., 2011). This can shift attention towards worries and consequences of not achieving the incentivized goal. The incentive does not necessarily have to be monetary; any manipulation that heightens the individual's metacognitive awareness of the performance situation might lead to rumination on potential outcomes, thereby reducing the attentional resources available for the task. This can effectively transform a single-task situation into a dual-task situation (Beilock & Carr, 2001; Beilock & Gray, 2007).

Outcome pressure's proposed method of skill failure is rooted in distraction theories (Beilock & Gray, 2007; Eysenck et al., 2007; Wine, 1971). According to Eysenck et al. (2007), anxiety increases the influence of the bottom-up stimulus-driven system at the expense of the top-down goal-driven system. This shift can cause several detrimental effects to task performance, especially when time-sharing between multiple tasks is needed. These effects include reduced inhibition of incorrect responses, increased susceptibility to distractions, and impaired task-switching performance. In a dynamic task environment where constant prioritization is needed, outcome pressure might cause attentional shifts more frequently towards less important targets. Therefore, we expect that in our task, performance will be lower and subtask prioritization will become less optimal under outcome pressure.

1.2.3 Previous studies comparing pressures

DeCaro et al. (2011) supported their theories of pressure classification by demonstrating in multiple experiments that outcome pressure induced distractions when tasks relied heavily on attentional control, while monitoring pressure impaired tasks that functioned best without conscious control. Endres et al. (2020) also found in separate experiments that outcome pressure worsened performances in an inhibition task requiring heavy attentional control, but monitoring pressure had no such effect.

In a rare within-subjects study comparing pressure categories by Soleimani Rad et al. (2022), participants experienced separate conditions for both monitoring and outcome pressure. Their task was to hit table tennis serves as accurately as possible while making quick decisions about the desired type of shot for each ball. The researchers found that decision-making accuracy was worse only under outcome pressure, whereas shot performance was worsened only by monitoring pressure. These results support the hypotheses laid out by DeCaro et al. (2011): cognitive decision-making task requiring attentional control was disrupted by outcome pressure, and the more proceduralized motor task of hitting table tennis balls was disrupted by monitoring pressure.

However, there is also contradictory evidence suggesting that monitoring pressure alone can worsen performance in a classic attentional control task (Belletier et al., 2015). This effect might be particularly pronounced in people with higher working memory capacity, possibly because they have the bandwidth to attend simultaneously to both the task and the presence of evaluative others. Similarly, there is evidence that outcome pressure alone can sometimes worsen performance in simple motor tasks (Geukes et al., 2013), which challenges the

hypotheses regarding monitoring and outcome pressure. Thus, more research is needed to better differentiate the effects of these two types of pressure. This is an underrepresented area of research in the field of choking under pressure (Soleimani Rad et al., 2022), and empirical within-subject studies comparing both types of pressures in the exact same task are particularly rare.

1.3 Goals of the study

Our study makes two main contributions. Firstly, we aim to contribute to the growing body of evidence showing that monitoring pressure and outcome pressure are distinct in how they affect performance and attentional processes, with each pressure type influencing individuals in different ways. This distinction is important for understanding the broader impact of pressure on cognitive and motor tasks. Secondly, we extend this pressure classification to a novel and demanding time-sharing environment, where the specific elements of pressure in multitasking contexts have not been thoroughly explored. By focusing on how time-sharing is affected under different types of pressure, our study provides new insights into how attentional resources are allocated across tasks.

Our experimental task contains four subtasks of varying priority, all requiring only the same visual attentional resources (see Section 2.2 for task description). The subtasks cannot be performed simultaneously; successful performance requires allocating attention and switching between them dynamically. Previous research on this task found that participants adapt to these varying priorities very quickly and successfully interact with the subtasks according to their importance (Kulomäki et al., 2022). This makes the task an excellent platform for examining the effects of pressure on time-sharing and prioritization.

Because the effects of impaired attentional control on performance can be compensated by modified and increased efforts (Eysenck et al., 2007), eye movements are often considered a better measure of attentional control (e.g., Luo et al., 2017; Wood & Wilson, 2010; Wright et al., 2014). Therefore, we recorded participants' eye movements in addition to task performance metrics to evaluate attention allocation.

1.3.1 Hypotheses

Our experimental task is designed to require heavy attentional control throughout performance, which should make it susceptible to outcome pressure-induced distractions. Given that the task requires conscious control and does not centrally involve proceduralized skills, it should not be significantly affected by monitoring pressure.

Hypothesis 1: Task performance scores will be lower under outcome pressure (compared to the no-pressure condition) and unaffected by monitoring pressure.

Hypothesis 2: Subtask prioritization will become less optimal under outcome pressure but will be unaffected by monitoring pressure. This will be evidenced in eye movements by higher event rate (i.e., more important) subtasks having less average time spent looking at them and them having lower visual sampling rate, compared to no-pressure condition. And vice versa, lower event rate subtasks will have more time and higher visual sampling rate under pressure.

2 Methods

The study was conducted in accordance with the Declaration of Helsinki and the participants were free to withdraw from the study at any time without consequences. The study and specifically its deceptive elements have been approved by the University of Turku's Ethics Committee for Human Sciences. To support open science, the study was preregistered on the Open Science Framework (OSF) prior to data collection

(https://osf.io/vrj7m/?view_only=72a7486270b948d3bbcd1d4815a4da0b) and all the data

used for analyses has been uploaded to our OSF Supplementary

(https://osf.io/g4bh5/?view_only=6f3c5fbd80354c21997c698021606c4e).

2.1 Participants

The target sample size was determined using an a priori power analysis with G*Power 3.1 (Faul et al., 2009). The results indicated that to achieve 80% power to detect a medium effect size ($f = .25$) with an alpha level of .05, a total of 28 participants would be needed. Due to our counterbalancing strategy, we required the sample size to be a multiple of six. Therefore, we aimed to recruit 30 participants.

The final sample contains 30 students from the University of Turku who volunteered for the experiment, which was conducted between November 2023 and March 2024. The average age of the 30 participants (24 females, 5 males, and 1 other) was 23.7 years, ranging from 20 to 37 years. Most of the participants reported being right-handed (26 right-handed and 4 left-handed). Participants were also asked about their average weekly video gaming habits: 17 reported playing no video games, 6 reported playing 1 to 5 hours weekly, 6 reported 6 to 10 hours weekly, and 1 reported playing over 20 hours weekly. Participants had the possibility to get both course credits and a 10€ gift voucher from the participation. Three experiment sessions were invalidated due to software malfunctions during the sessions and were replaced with new participants to maintain the target sample size of 30.

2.2 Stimuli

To evaluate participants' attentional control and performance under pressure, we used a computerized time-sharing task originally introduced and described in more detail, including a video sample, by Kulomäki et al. (2022). The task was inspired by the framework of instrument flying, where each instrument uniquely supports the task of flying an airplane.

However, the design was adapted to be generic and not limited to aviation. The subtasks were purposefully made simple to minimize the influence of experience, skill, and specialized knowledge. The task contained four overlapping subtasks of varying importance, requiring participants to dynamically switch between them. Participants were instructed to focus on all of the tasks without any information given about their priorities or possible strategies. This task paradigm was originally developed by Rantanen (Levinthal & Rantanen, 2004; Rantanen & Levinthal, 2005) and it has also been used successfully in experimental research more recently (M. A. Gray et al., 2023; Kulomäki et al., 2022). The computer task was created with E-Prime (Version 3.0).

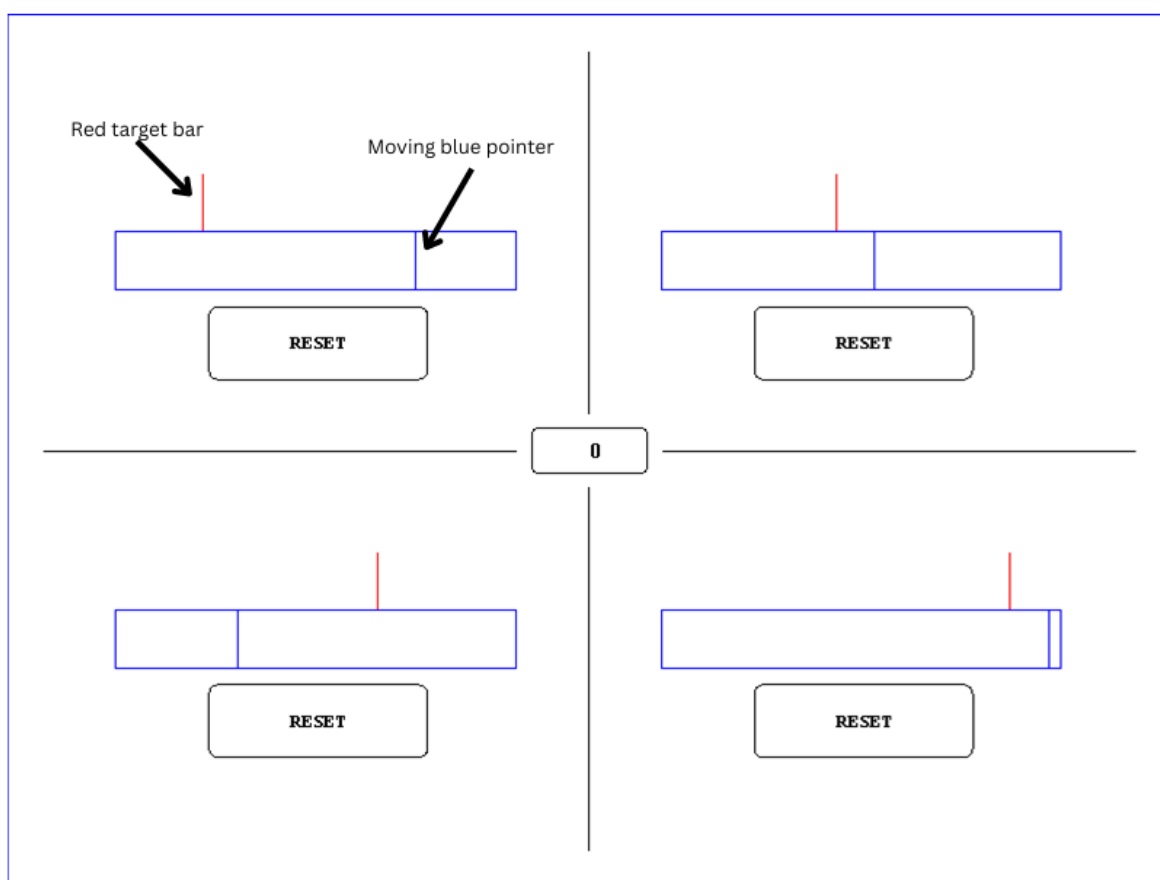


Figure 1. Screenshot from the computer task. The screen is divided into four subtasks of varying importances. The farther to the left the target bar is on the blue frame, the more important the subtask is, as the moving pointer meets with the target bar more frequently. In this figure, the subtask on the top left has the highest importance and the subtask on the lower right the lowest importance.

As shown in Figure 1, the screen was split into four quarters, each representing one subtask. Additionally, there was a score counter in the center. Each subtask contained a blue rectangular frame, with a moving blue pointer within it. Each frame also had a red target bar above it and a reset button below it. During the trials, the blue pointer automatically moved steadily from left to right, stopping only at the right edge of the frame. Participants could stop

the pointer and return it to the left edge by pressing the respective reset button with their mouse. The position of the red target bar varied between the four subtasks, determining the optimal frequency with which participants needed to interact with the subtask, i.e., the subtask event rate. The closer the red target bar is to the left edge of the frame, the faster and more often the moving pointer will meet it. The optimal frequencies which the pointer meets the target, or subtask event rates, were 0.34 Hz, 0.17 Hz, 0.11 Hz, and 0.08 Hz. Subtasks with higher event rate were more important for optimal performance and required more frequent interaction.

During the trials, the participants' task was to attend to all four subtasks and press the respective reset button as the blue pointer met with the respective red target bar. If the reset button was pressed while the pointer was horizontally within two pixels of the red target bar, the reset was considered successful. A successful reset was signaled by adding 10 points to the score counter in the middle of the screen and by playing a distinct high sound effect. If at the time of the reset the pointer was further than two pixels from the red target, a low sound effect was played, and two points were subtracted from the score counter. Additionally, if a pointer passed the respective target bar without the reset button being pressed, two points were subtracted, and a low sound was played once every second until the pointer was reset. The location of the subtasks with different event rates was randomized for each trial.

2.3 Measures

During the computer task, participants' eye movements were recorded using a desktop-mounted EyeLink 1000 Plus eye tracker. Movements of the right eye were tracked. The eye-movement data was analyzed to identify fixations, which were then assigned to five areas of interest: the four subtasks and the score counter. Main dependent measures were the average percentage of trial time spent looking at different areas of interests, visual sampling rate, and the composite score from the computer task. Visual sampling rate was defined as the total dwell time spent on a subtask divided by the number of gaze visits (enter and leave) to it.

To evaluate the effectiveness of our pressure manipulations, participants self-reported their perceived workload and feelings of anxiety during each condition using NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988) and 6-item Spielberger State-Trait Anxiety Inventory (STAI; Marteau & Bekker, 1992; Spielberger, 1970) forms, respectively. Both scales are widely used and validated measures with a long history of usage in experimental research (Hart, 2006; Rossi & Pourtois, 2012). In addition, we measured participants' pupil

size and compared it between conditions. Pupil size has been shown to be an effective objective measurement of mental stress: as stress or anxiety increases, so does the pupil size (e.g., Giannakakis et al., 2022; van der Wel & van Steenbergen, 2018; Yamanaka & Kawakami, 2009).

2.4 Design and Procedure

We manipulated one factor: the pressure condition, which had three levels (no-pressure, outcome pressure, and monitoring pressure). In a within-subjects design, each participant experienced all three conditions in a fully counterbalanced order.

First, all participants received verbal and written introductions to the study and provided written informed consent. Crucially, the participants were only informed that the study was about attention and attentional control, they were not aware of the upcoming pressure elements. Before the actual conditions, all participants completed four one-minute practice trials of the computer task to become accustomed to the task. Participants sat 62 cm from the computer screen and their heads were stabilized using a chinrest. The EyeLink camera was calibrated using a nine-point calibration before each condition, and drift correction was performed before each trial. The calibration error threshold was set at maximum of 1.0 degree for a single calibration point and a maximum average of 0.5 degrees for all the points. The lighting in the experiment room was kept constant across all conditions and participants to ensure accurate pupil measurements.

During each condition, after the experimenter gave the possible pressure-inducing instructions, participants filled out a 6-item STAI form (Marteau & Bekker, 1992) before starting the trials, to evaluate their current state anxiety. After completing all trials in a condition, they filled out a NASA-TLX form (Hart & Staveland, 1988), to report their perceived workload during the condition. After all the conditions, the participants were briefly interviewed about their strategies and whether they felt any kind of pressure during the tasks. Finally, everyone was fully debriefed and informed about the deception used during the experiment. The whole session lasted approximately 90 minutes.

All participants went through the following three conditions in a counterbalanced order.

2.4.1 No-pressure condition

Participant completed 10 one-minute trials of the computer task alone in the room. The experimenter left the room and could not see the participant or their computer screen during the trials. There was a 10-second break between trials.

2.4.2 Outcome pressure condition

Similarly to the no-pressure condition, the participant completed 10 one-minute trials alone in the room. However, as a deception to induce outcome pressure, the participant was informed before starting the trials that the researchers had predetermined an average score target that all participants should achieve during the experiment. They were told that their current average score in previous trials (practice and possible previous conditions) was lower than the target and thus potentially too low for the researchers to obtain high-quality data from the session.

The participant was encouraged to try improving their performance in the upcoming 10 trials, and that improving their score by 20% compared to their average score so far could earn them a €10 gift voucher. Additionally, the experimenter stated that the participants had been paired beforehand, and both the current participant and their pair would need to achieve the predetermined target for both to earn the gift voucher. The participant was told that their supposed pair had already completed the experiment and achieved the target, leaving it up to the present participant to improve for both individuals to be rewarded.

In reality, there was no pair, and all participants were eligible to receive compensation, regardless of their performance. During the condition, participants were not told their current average score or given a concrete score value to aim for, just that they needed to improve by around 20%. If they asked for the exact target, they were told that it was not revealed yet to encourage them to do their best, not just hit the minimum target. Similar deceptions have been shown to be effective in creating outcome pressure in previous literature (Beilock et al., 2004; DeCaro et al., 2011; Mullen et al., 2016; Smeding et al., 2015; Soleimani Rad et al., 2022).

2.4.3 Monitoring pressure condition

As in the other conditions, the participant completed 10 one-minute trials of the computer task. However, this time the experimenter was present in the room, filming the participant and

their computer screen with a video camera. The camera and experimenter were positioned behind and to the left of the participant, ensuring that the participant's computer screen was fully visible to the experimenter. The participant was informed that both they and their screen were being filmed and observed during the task.

Participant was informed that the video might be viewed by other researchers involved in the study to evaluate their performance and that it could additionally be used as presentation material in a future course about psychology lab work. In reality, the camera was only used to induce pressure, and all the video material was immediately deleted without being viewed by anyone. Similar instructions have been shown to be effective at inducing monitoring pressure in previous literature (eg., DeCaro et al., 2011; Endres et al., 2020; Mesagno et al., 2011; Smeding et al., 2015; Soleimani Rad et al., 2022).

3 Results

In this section, we first report the subjective and objective measures of the workload and pressure experienced by the participants during the experiment. Next, we present the results of our confirmatory analyses in line with [our preregistration](#). This is followed by additional in-depth exploratory analyses. Assumptions were adequately met for all reported methods. Analyses were conducted with IBM SPSS Statistics (Version 29). All the data used in our analyses is freely available on our online [OSF Supplementary](#).

3.1 Pressure manipulation

To evaluate the effectiveness of our pressure manipulations, we compared the 6-item STAI scores, unweighted NASA-TLX total scores, and pupil size between the no-pressure and pressure conditions. The means and standard deviations are reported in Table 1. In all three instruments, the no-pressure condition had lower scores than both pressure conditions, possibly indicating that participants felt higher pressure and anxiety in the manipulated conditions. However, in paired t-tests, only the differences in pupil size values were statistically significant at the $p < .05$ level (no-pressure vs. monitoring: $t(29) = -2.27, p = .031, d = -0.414$; no-pressure vs. outcome: $t(29) = -3.18, p = .008, d = -0.58$), with a medium effect size. Neither the paired t-tests for STAI (no-pressure vs. monitoring: $t(29) = -0.63, p = .534, d = -0.12$; no-pressure vs. outcome: $t(29) = -1.95, p = .120, d = -0.36$) nor NASA-TLX reached statistical significance (no-pressure vs. monitoring: $t(29) = -0.24, p = .815, d = -0.43$; no-pressure vs. outcome: $t(29) = -1.03, p = .620, d = -0.19$). Reported p-values are adjusted for multiple comparisons using the Holm-Bonferroni method (Holm, 1979).

Table 1.

Table containing descriptive statistics of measures for evaluating pressure participants experienced in each experimental condition.

Measure	No-pressure		Monitoring pressure		Outcome pressure	
	<i>M (SD)</i>	95% CI	<i>M (SD)</i>	95% CI	<i>M (SD)</i>	95% CI
State anxiety (six-item STAI)	11.50 (2.00)	[10.76, 12.25]	11.73 (2.61)	[10.76, 12.71]	12.40 (3.07)	[11.25, 13.55]
Perceived workload (NASA-TLX)	59.94 (14.50)	[54.53, 65.36]	60.33 (14.94)	[54.75, 65.91]	61.58 (14.73)	[56.08, 67.08]
Pupil size (EyeLink-measured number of pixels)	1016.11 (270.56)	[915.08, 1117.14]	1045.82 (259.81)	[948.80, 1142.83]	1061.15 (279.36)	[956.84, 1165.46]

3.2 Confirmatory Analysis

The following confirmatory analyses were conducted as specified in [our preregistration](#). To analyze the effects of pressure on participants' composite scores in the computer task (Hypothesis 1), we conducted a one-way repeated measures ANOVA with the composite score as the dependent variable and pressure situation as the independent variable with three levels (no-pressure, monitoring pressure, and outcome pressure). To analyze the effects of pressure on participants' subtask prioritization (Hypothesis 2), we conducted separate repeated measures ANOVAs with the same three pressure levels for two dependent variables: average percentage of trial time spent looking at each subtask and visual sampling rate of each subtask. For both variables, the ANOVAs were repeated for each subtask (0.08 Hz, 0.11 Hz, 0.17 Hz, and 0.34 Hz), resulting in a total of eight separate models.

Relatively simple repeated measures ANOVAs were chosen to maximize the statistical power given the sample size available for this study. While linear mixed-effect models (LMMs) could provide a more flexible approach by accounting for both fixed and random effects, they typically require larger sample sizes to maintain similar statistical power. The increased complexity of LMMs would introduce more parameters and thus increase the risk of overfitting with a small sample, potentially reducing the reliability of the results. Therefore, repeated measures ANOVAs were preferred for their balance of simplicity and efficiency in this context.

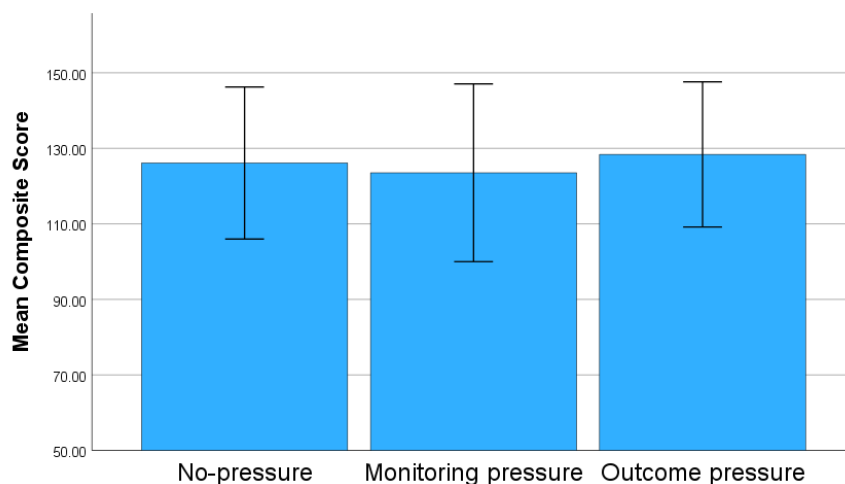


Figure 2. Mean composite scores in different pressure conditions. Error bars represent 95 % confidence intervals.

The composite score values in different pressure conditions are visualized in Figure 2. The repeated measures ANOVA revealed no significant differences in score values between pressure conditions ($F(2, 58) = 0.17, p = .845, \eta_p^2 = 0.01$). This indicates that being under either kind of pressure did not affect the score values (no-pressure 126.1, monitoring pressure 123.5, and outcome pressure 128.4).

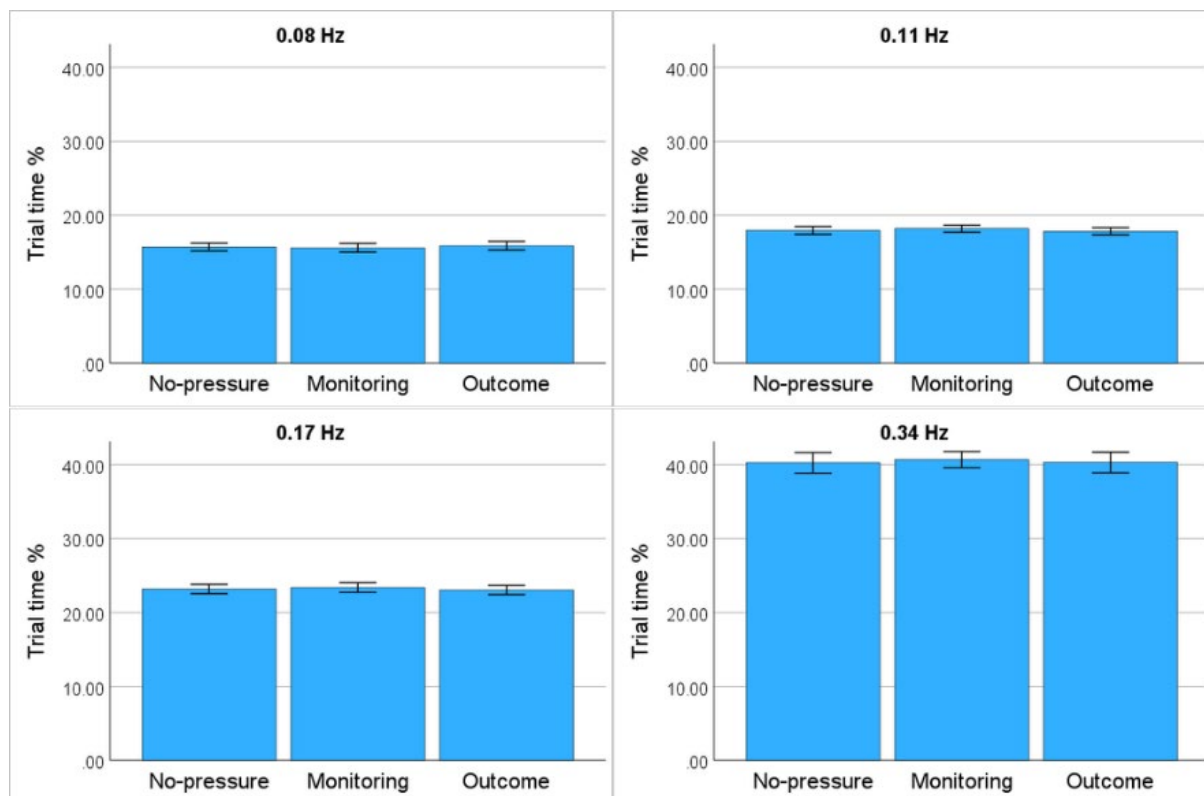


Figure 3. Means of the percentage of time spent looking at each subtask during the trials. The four subtasks are presented in separate panes. Error bars represent the 95% confidence intervals.

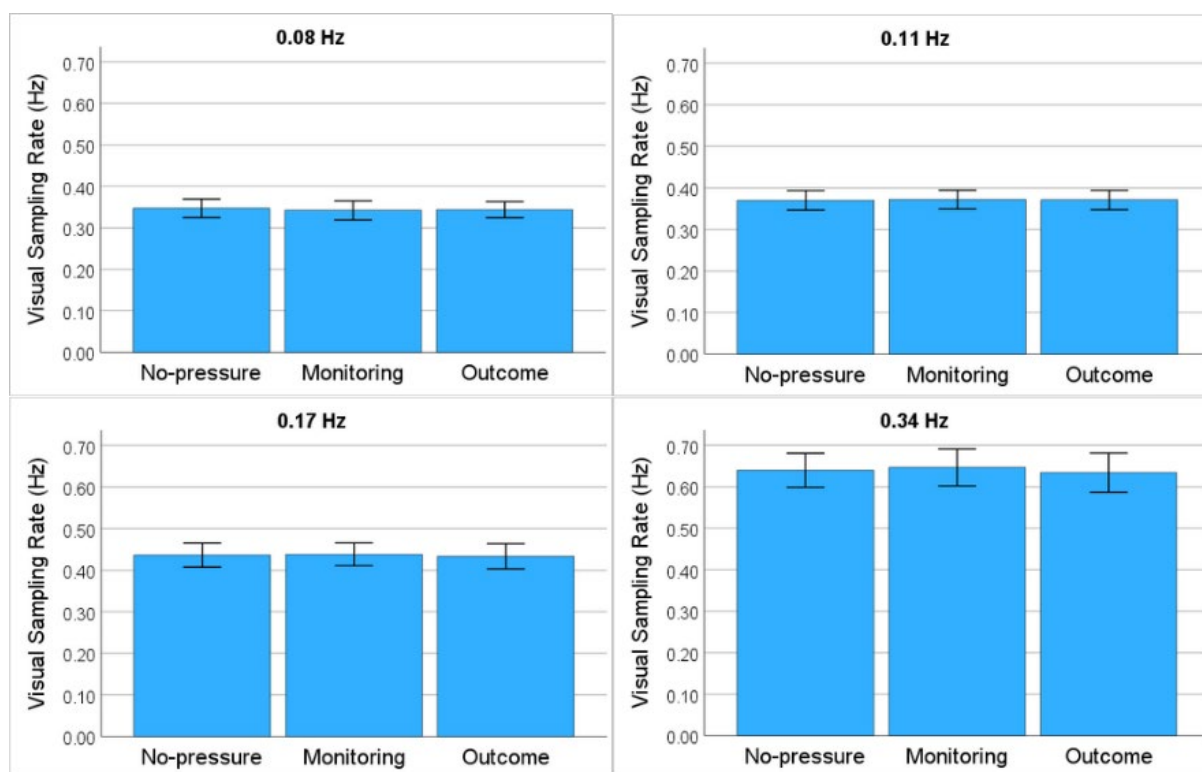


Figure 4. Means of the visual sampling rate in each of the subtasks, which are presented in separate panes. Error bars represent the 95% confidence intervals.

The percentage of time participants spent looking at each subtask and the visual sampling rates of the subtasks are visualized for each pressure situation in Figures 3 and 4. The full results of the eight repeated measures ANOVAs are presented in Table 2. None of the models presented statistically significant results. This suggests that participants' ability to prioritize subtasks seems to have been unaffected by the experienced pressure and they allocated attention between the subtasks similarly across all conditions.

Table 2.

Results from the eight repeated measures ANOVAs conducted for the two dependent variables, separately for each subtask, which are identified by their event rates.

Dependent Variable	Subtask	$F(2,58)$	p	η_p^2
Percentage Trial Time Spent	0.08 Hz	0.60	.552	0.02
	0.11 Hz	1.04	.359	0.04
	0.17 Hz	0.42	.657	0.01
	0.34 Hz	0.38	.685	0.01
Visual Sampling Rate	0.08 Hz	0.56	.572	0.02
	0.11 Hz	0.04	.958	0.01
	0.17 Hz	0.21	.812	0.01
	0.34 Hz	0.77	.467	0.03

3.3 Exploratory Analysis

In this section, we present our exploratory analyses conducted in addition to the hypothesis testing. The primary goal here is to explore data, uncover patterns, and generate new insights rather than confirm specific hypotheses. Consequently, the p-values reported in this section are unadjusted to avoid false negatives, accepting the heightened risk of false positives.

Participants exhibited statistically significant differences in pupil sizes under pressure situations compared to no-pressure conditions. However, in post-experiment interviews, some participants admitted they were somewhat skeptical of the instructions intended to induce pressure. Although none expressed complete disbelief, it is possible that some participants did not feel pressured at all by the manipulations. These individual differences in perceived pressure are also evident in Figure 5, which shows the average differences in pupil sizes between pressure and no-pressure conditions for each participant. The difference in pupil size for each participant was calculated by taking the average change from the no-pressure condition to the two pressure conditions. Specifically, this was computed as the mean of the differences between each pressure condition and the no-pressure condition:

$$\text{Average Difference} = \frac{(\text{Monitoring} - \text{No-pressure}) + (\text{Outcome} - \text{No-pressure})}{2}$$

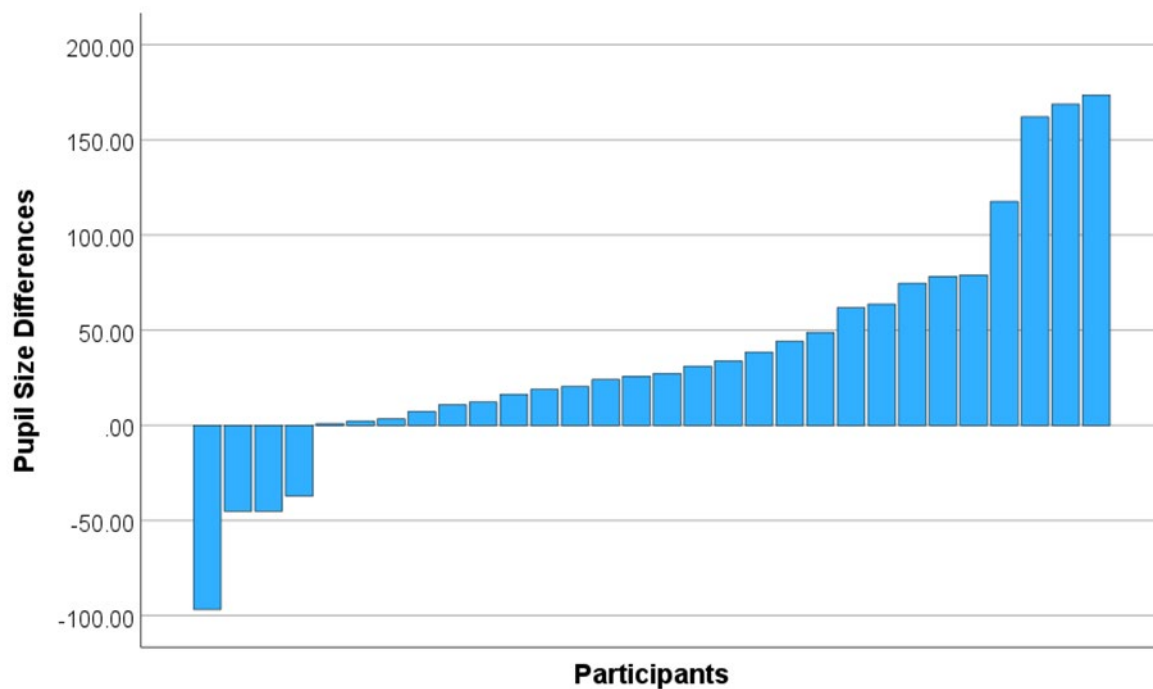


Figure 5. Visualized average changes for each participant's pupil size while under pressure. Higher positive values indicate that the average pupil size was larger while under pressure.

Due to these individual differences, we proceeded to split the dataset in half based on the calculated pupil size differences. We then conducted additional analyses on the half with the highest positive pupil size differences, i.e., the participants for whom the manipulations appeared to be the most effective. Pupil sizes in this high-pressure subset were similar to the complete dataset, with no-pressure condition having the lowest mean (947.24), monitoring pressure being second highest (1022.43), and finally outcome pressure as the highest (1032.49). Differences in pupil size between no-pressure and both pressure conditions were statistically significant (monitoring: $t(14) = -4.24, p < .001, d = -1.10$; outcome: $t(14) = -4.08, p < .001, d = -1.05$), but the differences between monitoring and outcome pressure were not significant ($t(14) = -0.36, p = .728, d = -0.09$).

To reanalyze the composite score values under pressure, we repeated the previously reported one-way repeated measures ANOVA for the high-pressure subset of the dataset ($n = 15$). This time, the effect of pressure on task score was statistically significant ($F(2, 28) = 3.53, p = .043, \eta_p^2 = 0.20$). Planned contrasts revealed that the average score under monitoring pressure (108.76) was statistically significantly lower than the score under no-pressure (135.78), with a medium effect size ($t(14) = 2.77, p = .015, d = 0.72$). However, the difference in scores between outcome pressure (127.48) and no-pressure (135.78) was not statistically significant ($t(14) = 0.80, p = .438, d = 0.21$). Analyses for the average trial time spent looking at different subtasks and visual sampling rates were also repeated on this high-pressure subset of the data, but the results (reported in Table 3) did not differ from those previously reported for the complete dataset.

Table 3.

Additional analyses conducted for the high-pressure half of the data. Results from the eight repeated measures ANOVAs conducted for time spent looking at each subtask and visual sampling rate, reported separately for each subtask.

Dependent Variable	Subtask	$F(2,28)$	p	η_p^2
Percentage Trial Time Spent	0.08 Hz	0.36	.700	0.03
	0.11 Hz	3.14	.059	0.18
	0.17 Hz	1.96	.160	0.12
	0.34 Hz	0.15	.863	0.01
Visual Sampling Rate	0.08 Hz	0.83	.447	0.06
	0.11 Hz	1.00	.381	0.07
	0.17 Hz	0.01	.995	0.01
	0.34 Hz	0.05	.954	0.01

To investigate what might have caused the lower scores while under monitoring pressure, we then explored the reset errors made by participants, i.e., the instances when they pressed the reset button while the corresponding pointer was over two pixels away from the corresponding target bar. All average reset errors for each experimental condition are presented in Table 4, separately for the whole data and for the high-pressure subset. Repeated measures ANOVA conducted on the complete dataset revealed no significant effect of pressure for the reset error ($F(2, 58) = 0.17, p = .841, \eta_p^2 = 0.01$). For the high-pressure subset, the reset errors were on average over 1 pixel higher for both pressure conditions, when compared to the no-pressure condition. Meaning, participants made slightly larger errors while under both pressures. However, the effect of pressure in repeated measures ANOVA for the high-pressure subset was not statistically significant ($F(2, 28) = 1.95, p = .161, \eta_p^2 = 0.12$).

Table 4.

Average reset errors in different conditions, separately for the whole data and high-pressure half of the data.

Condition	Reset error in pixels (SD)	
	All data	High-pressure half
No-pressure	8.21 (3.07)	7.47 (2.04)
Monitoring pressure	7.95 (3.22)	8.58 (2.90)
Outcome pressure	8.21 (3.02)	8.49 (2.94)

Next, we classified the reset errors further as early errors and late errors. Early errors were defined as reset button presses happening while the moving pointer was more than two pixels to the left of the target bar. Late errors happened when the moving pointer was more than two pixels to the right of the target bar. The number of early and late errors in different conditions are reported in Table 5. Across all conditions and in both the complete data and the high-pressure subset of the data, participants made around 4 to 6 times more late errors than early errors. For the complete data, values for both early and late errors are similar across all conditions. For the high-pressure subset, late errors are similar in all conditions. However, the early errors during monitoring pressure stand out as higher than other conditions in the high-pressure subset. Repeated measures ANOVA indicates that for the high-pressure subset the

effect of pressure on the number of early errors is not statistically significant, but very close to it ($F(2, 28) = 3.25, p = .054, \eta_p^2 = 0.19$). A post hoc paired t-test also indicates that the number of early errors in monitoring pressure is significantly higher compared to no-pressure ($t(14) = 2.41, p = .030, d = 0.62$).

Table 5.

The number of early and late reset errors in different pressure conditions, separately for the whole data and high-pressure half of the data.

Condition	All data		High-pressure half	
	Early errors	Late errors	Early errors	Late errors
No-pressure	35.33 (27.53)	155.63 (39.18)	27.67 (16.58)	156.47 (40.38)
Monitoring pressure	35.83 (25.92)	153.40 (35.02)	40.33 (24.67)	160.13 (23.61)
Outcome pressure	37.50 (31.71)	153.10 (32.26)	29.93 (16.94)	159.40 (36.08)

4 Discussion

The aim of our study was to investigate whether monitoring pressure and outcome pressure exert differing effects on time-sharing performance and allocation of attention. Our first hypothesis predicted that, when compared to the no-pressure condition, outcome pressure would lower task performance scores, while monitoring pressure would have no significant effect. Similarly, our second hypothesis stated that subtask prioritization would deteriorate under outcome pressure but remain unaffected by monitoring pressure. However, our pre-registered confirmatory analyses provided no evidence to support either of these hypotheses.

The mean task scores were consistent across all three conditions, with only minor and not statistically significant variations that likely resulted from random chance. A similar pattern emerged for subtask prioritization, as measured by eye movements. Participants allocated their attention and time similarly among the different subtasks across conditions. Consistent with previous research using the same task (Kulomäki et al., 2022), participants appropriately allocated the majority of their time (approximately 40%) to the most important subtask, while dedicating considerably less time (15-23%) to each of the other three subtasks. The visual sampling rate further supports these findings, showing no significant differences in how participants sampled the subtasks, regardless of the experimental condition. Notably, there was no evidence of attentional tunneling towards the most important subtask(s) under pressure. One possible explanation for the absence of focused attention on select subtask(s) under pressure could be the sound effects used in our study. These sound cues alerted participants every second they failed to interact with a subtask, potentially preventing the development of attentional tunneling. It is well-documented that salient alarms for neglected tasks can mitigate attentional tunneling (C. Wickens, 2021).

To evaluate the effectiveness of our pressure manipulations, we used both subjective self-report measures and an objective physiological measure: pupil size. While the mean scores for self-reported workload (NASA-TLX) and state anxiety (STAI) were higher in the pressure conditions compared to the no-pressure condition, these differences were not statistically significant. Notably, even in the no-pressure condition, NASA-TLX scores were fairly high, averaging 60 points. Previous meta-analyses have found the mean recorded NASA-TLX scores across hundreds of studies to range between 42 and 49 (Grier, 2015; Hertzum, 2021). For instance, Grier (2015) reported a median score of 46.00 across 31 cognitive task studies, 52.24 across 174 monitoring tasks in different studies, and 52.44 in 24 tasks involving air

traffic control. Our participants consistently reported higher values, even in the no-pressure condition. This suggests that our task may be inherently very demanding, potentially limiting the ability of subjective measures like NASA-TLX and STAI to effectively distinguish between our pressure and no-pressure conditions. As the workload and experienced pressure increases towards the extreme ranges, it may become more challenging for participants to subjectively differentiate between the already elevated values.

In contrast to the subjective measures, the differences in pupil size between the no-pressure and pressure conditions were statistically significant, indicating that our manipulations were generally effective. Because pupil size is very sensitive to lighting changes in the environment, we made sure that the lighting was kept as constant as possible across conditions. The only changes between conditions in the room was the presence of the experimenter and camera during monitoring pressure and those did not significantly alter the lighting environment. The experimental task was identical in all conditions.

However, a few participants admitted after the sessions that they did not necessarily fully believe the instructions intended to induce pressure. Although they generally still indicated that they were unsure whether the instructions were false and believed that they were still likely somewhat affected by them, we chose to further analyze the data with this in mind. We split the dataset in half based on the degree of change in pupil sizes between the pressure and no-pressure conditions. We then conducted additional exploratory analyses on the high-pressure half of the dataset, i.e. those who, according to their pupil responses, reacted most strongly to the pressure manipulations.

Repeating the previous analyses on the high-pressure subset of data revealed that participants' task scores were statistically significantly lower while under monitoring pressure compared to the no-pressure condition. While scores were also slightly lower under outcome pressure, the differences were not statistically significant. These findings seem to contradict the prevailing theories of monitoring and outcome pressure, which suggest that our task should be more susceptible to the distractions caused by outcome pressure rather than the negative effects typically associated with monitoring pressure. However, there is existing evidence that monitoring pressure can also lead to distractions (Belletier et al., 2015). Importantly, these findings cannot be attributed to this subset of participants simply feeling more pressured under monitoring pressure, as outcome pressure actually resulted in higher pupil size than monitoring pressure in the high-pressure subset, though this difference was non-significant.

To explore the reason behind the lower scores under monitoring pressure, we analyzed the specific errors participants made during the task. In our task, each subtask involved a pointer that moved slowly from left to right. Participants had to press a reset button to return the pointer to the left edge of the subtask when the pointer was aligned with a corresponding target bar. They were allowed leeway of two pixels on either side of the target bar. If participants pressed the reset button when the pointer was more than two pixels to the left of the target bar, they made an early error and lost points. Conversely, if they pressed the reset button when the pointer was more than two pixels to the right of the target bar, they made a late error and lost points.

We first examined the overall magnitudes of reset errors across each condition. In the complete dataset, as expected, the errors were similar across conditions, with only minor random variations. In the high-pressure subset, errors were consistently higher by approximately one pixel in both pressure conditions compared to the no-pressure condition. However, these differences were not statistically significant, and since the error magnitudes were similar under both outcome and monitoring pressure, they could not account for the previously observed differences in task scores between pressure conditions.

Next, we analyzed the frequency of early and late reset errors. In the complete dataset, no significant differences were observed. However, for the high-pressure subset, we found a statistically significant increase in the number of early errors under monitoring pressure compared to the no-pressure condition. No significant differences were found for outcome pressure or late errors. This increase in early errors appears to be the key factor explaining the lower task scores under monitoring pressure. Since there were no differences in the magnitude of errors between pressure conditions, this suggests that participants made slightly smaller errors but did so more frequently under monitoring pressure, leading to a lower overall task score.

A significant portion of explicit monitoring research supporting the theory of monitoring pressure focuses on proceduralized sensorimotor skills, such as golf putting (Beilock & Carr, 2001), baseball batting (R. Gray, 2004), and table tennis (Soleimani Rad et al., 2022). We did not anticipate that monitoring pressure would impair performance in our task, which relies heavily on attentional control rather than proceduralized skills. However, it can be argued that our task does, in fact, involve a significant sensorimotor component: the precise and rapid use of the mouse. The task is highly time-sensitive, requiring quick and accurate mouse usage.

Being monitored by the experimenter and recorded by the video camera may have made participants feel that their performance was being closely scrutinized, leading them to become more self-conscious about their mouse movements and clicks. This heightened self-awareness might have caused them to rush their button presses, resulting in early errors and decreased accuracy.

Therefore, we suggest that monitoring pressure impaired task performance not by distracting participants from the task itself but by disrupting the motor task of accurately using the mouse. This aligns with the expected effects of monitoring pressure on sensorimotor tasks and is supported by the fact that no significant changes are observable in the eye-tracking data representing attentional allocation. Additionally, it underscores the importance of carefully planning experimental tasks in future research. If monitoring pressure can interfere with the precise use of instruments like a mouse, which is commonly used in computer-based tasks, laboratory experiments investigating psychological pressure must take this factor into account.

However, alternative explanations are also worth considering, since we did not directly measure sensorimotor disruption, and our explanation relies on inferential reasoning. It is also possible that under the pressure of being observed, participants might have been overly focused on their performance and worried about being judged, diverting cognitive resources away from the task, a mechanism more commonly associated with outcome pressure. This cognitive overload could indirectly impair task performance by making it harder to manage on the dynamic and complex task demands, leading to poor decision-making or timing errors. Alternatively, participants might have wanted to give a good impression while being observed and recorded, resulting in faster reactions at the cost of accuracy.

Outcome pressure, while not yielding significant results in this study, still probably plays a role in performance failure under different cognitive demands or even with time-sharing performance. Future research should focus on directly measuring the cognitive and motor mechanisms behind performance under both types of pressure. This could include real-time assessments of e.g., motor precision, potential rumination under outcome pressure behind the distraction, cognitive load, and physiological markers of stress (e.g., heart rate variability, electrodermal activity, pupil dilation) to better understand the distinct and potentially interactive effects of monitoring and outcome pressures on task performance.

4.1 Study Limitations

Our study has certain limitations that should be acknowledged for consideration in future research. Firstly, while we argue that our pressure manipulations were effective, as evidenced by changes in pupil size, there might be room for further refinement. For outcome pressure, instead of a general request to improve performance, such as the 20% improvement we employed without providing the actual target score, some studies have used specific score targets that participants must achieve (e.g., Mullen et al., 2016). It is possible that a concrete target could increase motivation and introduce more distractions due to the heightened pressure. However, approach carries potential risks: if the target is set too high, some participants might become discouraged and give up, whereas those who reach the target early might relax for the remainder of the session. Additionally, offering more valuable rewards or enhancing the social pressure aspect, such as by having participants actually see someone they believe is their partner, could potentially increase the effectiveness of the pressure manipulation.

For monitoring pressure, some studies have introduced an additional person into the room during the pressure condition, whose sole purpose is to evaluate the participants performance (Soleimani Rad et al., 2022). This may exert more pressure on the participant compared to just the presence of the main experimenter and cameras. It should be noted, however, that most other recent pressure studies have used manipulations identical or very similar to ours, with successful results (e.g., DeCaro et al., 2011; Endres et al., 2020; Smeding et al., 2015).

Additionally, the pressure experienced by participants could be measured even more precisely by incorporating further physiological measures known for their accuracy in this domain, such as heart rate variability (Kim et al., 2018) or electrodermal activity (Giannakakis et al., 2022). However, using more physiological sensors comes with the tradeoff of increased invasiveness, which could introduce additional stress and anxiety due to the presence of the measuring equipment.

Individual differences among participants could be more thoroughly accounted for in future research. There is evidence suggesting that factors such as working memory capacity can influence who is more likely to be distracted under pressure (Beilock & Carr, 2005; Belletier et al., 2015). Additionally, individuals vary in their predisposition to stress, particularly in terms of trait anxiety levels (Eysenck, 1992; Eysenck et al., 2007). In this study, there was an uneven gender distribution, with 24 participants identifying as female, 5 as male, and 1 as

other. While gender was not a specific variable of interest in this experiment, we acknowledge that this imbalance could be a limitation. However, given the nature of the task and the primary focus on psychological pressure and time-sharing performance, we do not expect the gender distribution to have significantly influenced our key findings. Research on gender differences in performance under pressure yields mixed results. Some studies suggest that women (e.g., Cai et al., 2019) may be more negatively affected by pressure, while others show the same for men (e.g., Bühren et al., 2024). Additionally, some research reports no significant differences between genders (e.g., Ariely et al., 2009). Future studies could benefit from a more balanced sample to further investigate whether gender plays a role in pressure-induced performance effects. Even if these individual differences are not the primary focus of a study, not controlling for them may sometimes impact the results.

Some previous research has attempted to distinguish time pressure from other types of pressure (Endres et al., 2020). Our task inherently includes tight time constraints, as is typical in time-sharing environments. If time pressure has distinct effects or significantly interacts with other forms of pressure, its influence might be obscured when combined with other pressure types. Future research should consider this and potentially investigate time pressure as a separate, third type of pressure.

It is also important to acknowledge that some of our findings were derived from exploratory analyses of a limited sample size. As such, these results should not be viewed as definitive evidence. Instead, these exploratory findings should serve as a foundation for generating new hypotheses to be tested in subsequent research.

4.2 Conclusions

This study aimed to explore the differential effects of monitoring and outcome pressure on time-sharing performance and attentional allocation. While our findings did not support the initial hypotheses that outcome pressure would impair performance more than monitoring pressure, additional exploratory analyses revealed an unexpected impact of monitoring pressure, potentially attributable to the sensorimotor demands of the task. These results suggest that even in tasks primarily reliant on attentional control, the motor components, such as precise mouse usage, could potentially be disrupted by monitoring pressure. This highlights the complexity of psychological pressure and underscores the importance of considering task-specific factors when designing experiments and interpreting results. Performance pressure, such as mere experimenter presence, could also unintentionally influence e.g., computer-

based tasks in experiments not specifically designed to study or account for pressure.

Therefore, researchers should be mindful of the potential for such unintended influences when setting up experimental environments, particularly in studies where precise motor control or attentional focus is critical.

References

- APA dictionary of psychology*. (n.d.). Retrieved July 19, 2024, from <https://dictionary.apa.org/>
- Ariely, D., Gneezy, U., Loewenstein, G., & Mazar, N. (2009). Large stakes and big mistakes. *Review of Economic Studies*, *76*(2), 451–469. <https://doi.org/10.1111/j.1467-937X.2009.00534.x>
- Baddeley, A. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology Section A*, *49*(1), 5–28. <https://doi.org/10.1080/713755608>
- Baumeister, R. F. (1984). Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology*, *46*(3), 610–620. <https://doi.org/10.1037/0022-3514.46.3.610>
- Beck, A. T., & Clark, D. A. (1997). An information processing model of anxiety: Automatic and strategic processes. *Behaviour Research and Therapy*, *35*(1), 49–58. [https://doi.org/10.1016/s0005-7967\(96\)00069-1](https://doi.org/10.1016/s0005-7967(96)00069-1)
- Beilock, S. L., & Carr, T. H. (2001). On the fragility of skilled performance: What governs choking under pressure? *Journal of Experimental Psychology: General*, *130*(4), 701–725. <https://doi.org/10.1037/0096-3445.130.4.701>
- Beilock, S. L., & Carr, T. H. (2005). When high-powered people fail: Working memory and “choking under pressure” in math. *Psychological Science*, *16*(2), 101–105.
- Beilock, S. L., & Gray, R. (2007). Why do athletes choke under pressure? In *Handbook of sport psychology, 3rd ed* (pp. 425–444). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118270011.ch19>
- Beilock, S. L., Kulp, C. A., Holt, L. E., & Carr, T. H. (2004). More on the fragility of performance: Choking under pressure in mathematical problem solving. *Journal of Experimental Psychology: General*, *133*(4), 584–600. <https://doi.org/10.1037/0096-3445.133.4.584>
- Belletier, C., Davranche, K., Tellier, I. S., Dumas, F., Vidal, F., Hasbroucq, T., & Huguet, P. (2015). Choking under monitoring pressure: Being watched by the experimenter reduces executive attention. *Psychonomic Bulletin & Review*, *22*(5), 1410–1416. <https://doi.org/10.3758/s13423-015-0804-9>

- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652.
<https://doi.org/10.1037/0033-295X.108.3.624>
- Bühren, C., Gschwend, M., & Krumer, A. (2024). Expectations, gender, and choking under pressure: Evidence from alpine skiing. *Journal of Economic Psychology*, *100*, 102692.
<https://doi.org/10.1016/j.joep.2023.102692>
- Burgoyne, A. P., & Engle, R. W. (2020). Attention control: A cornerstone of higher-order cognition. *Current Directions in Psychological Science*, *29*(6), 624–630.
<https://doi.org/10.1177/0963721420969371>
- Cai, X., Lu, Y., Pan, J., & Zhong, S. (2019). Gender gap under pressure: Evidence from China's national college entrance examination. *Review of Economics & Statistics*, *101*(2), 1–15.
https://doi.org/10.1162/rest_a_00749
- DeCaro, M. S., Thomas, R. D., Albert, N. B., & Beilock, S. L. (2011). Choking under pressure: Multiple routes to skill failure. *Journal of Experimental Psychology: General*, *140*, 390–406.
<https://doi.org/10.1037/a0023466>
- Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, *66*(3), 183–201. <https://doi.org/10.1037/h0047707>
- Endres, D. N., Byrne, K. A., Anaraky, R. G., Adesegun, N., Six, S. G., & Tibbett, T. P. (2020). Stop the clock because I can't stop: Time pressure, but not monitoring pressure, impairs response inhibition performance. *Journal of Cognitive Psychology*, *32*(7), 627–644.
<https://doi.org/10.1080/20445911.2020.1810692>
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19–23. <https://doi.org/10.1111/1467-8721.00160>
- Eysenck, M. W. (1992). *Anxiety: The cognitive perspective* (pp. xii, 195). Lawrence Erlbaum Associates, Inc.
- Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, *7*, 336–353. <https://doi.org/10.1037/1528-3542.7.2.336>

- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Geukes, K., Mesagno, C., Hanrahan, S. J., & Kellmann, M. (2013). Performing under pressure in private: Activation of self-focus traits. *International Journal of Sport and Exercise Psychology*, *11*(1), 11–23. <https://doi.org/10.1080/1612197X.2012.724195>
- Giannakakis, G., Grigoriadis, D., Giannakaki, K., Simantiraki, O., Roniotis, A., & Tsiknakis, M. (2022). Review on psychological stress detection using biosignals. *IEEE Transactions on Affective Computing*, *13*(1), 440–460. *IEEE Transactions on Affective Computing*. <https://doi.org/10.1109/TAFFC.2019.2927337>
- Gray, M. A., Yong, Z., Wasti, A., Rantanen, E. M., & Heard, J. R. (2023). Measuring temporal awareness for human-aware AI. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *67*(1), 1817–1823. <https://doi.org/10.1177/21695067231192635>
- Gray, R. (2004). Attending to the execution of a complex sensorimotor skill: Expertise differences, choking, and slumps. *Journal of Experimental Psychology. Applied*, *10*(1), 42–54. <https://doi.org/10.1037/1076-898X.10.1.42>
- Grier, R. A. (2015). How high is high? A meta-analysis of NASA-TLX global workload scores. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 1727–1731. <https://doi.org/10.1177/1541931215591373>
- Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904–908. <https://doi.org/10.1177/154193120605000909>
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology* (Vol. 52, pp. 139–183). North-Holland. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)

- Hertzum, M. (2021). Reference values and subscale patterns for the task load index (TLX): A meta-analytic review. *Ergonomics*, *64*(7), 869–878.
<https://doi.org/10.1080/00140139.2021.1876927>
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.
- Jackson, R. C., Ashford, K. J., & Norsworthy, G. (2006). Attentional focus, dispositional reinvestment, and skilled motor performance under pressure. *Journal of Sport & Exercise Psychology*, *28*(1), 49–68. <https://doi.org/10.1123/jsep.28.1.49>
- Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y. H., & Koo, B.-H. (2018). Stress and heart rate variability: A meta-analysis and review of the literature. *Psychiatry Investigation*, *15*(3), 235–245. <https://doi.org/10.30773/pi.2017.08.17>
- Kulomäki, J., Oksama, L., Rantanen, E., & Hyönä, J. (2022). Attention control in a demanding dynamic time-sharing environment: An eye-tracking study. *Attention, Perception, & Psychophysics*, *84*(2), 352–371. <https://doi.org/10.3758/s13414-021-02377-z>
- Levinthal, B. R., & Rantanen, E. M. (2004). Measurement of taskload and performance in a dynamic multi-task experiment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(3), 567–570. <https://doi.org/10.1177/154193120404800362>
- Lewis, B. P., & Linder, D. E. (1997). Thinking about choking? Attentional processes and paradoxical performance. *Personality and Social Psychology Bulletin*, *23*(9), 937–944.
<https://doi.org/10.1177/0146167297239003>
- Luo, X., Zhang, L., & Wang, J. (2017). The benefits of working memory capacity on attentional control under pressure. *Frontiers in Psychology*, *8*.
<https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01105>
- Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State—Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology*, *31*(3), 301–306. <https://doi.org/10.1111/j.2044-8260.1992.tb00997.x>
- Masters, R., & Maxwell, J. (2008). The theory of reinvestment. *International Review of Sport and Exercise Psychology*, *1*(2), 160–183. <https://doi.org/10.1080/17509840802287218>

- Masters, R. S. W. (1992). Knowledge, knerves and know-how: The role of explicit versus implicit knowledge in the breakdown of a complex motor skill under pressure. *British Journal of Psychology*, 83(3), 343–358. <https://doi.org/10.1111/j.2044-8295.1992.tb02446.x>
- Mesagno, C., Harvey, J. T., & Janelle, C. M. (2011). Self-presentation origins of choking: Evidence from separate pressure manipulations. *Journal of Sport and Exercise Psychology*, 33(3), 441–459. <https://doi.org/10.1123/jsep.33.3.441>
- Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part I. Basic mechanisms. *Psychological Review*, 104(1), 3–65. <https://doi.org/10.1037/0033-295X.104.1.3>
- Mullen, R., Jones, E. S., Oliver, S., & Hardy, L. (2016). Anxiety and motor performance: More evidence for the effectiveness of holistic process goals as a solution to the process goal paradox. *Psychology of Sport and Exercise*, 27, 142–149. <https://doi.org/10.1016/j.psychsport.2016.08.009>
- Norman, D. A., & Shallice, T. (1986). Attention to action. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and Self-Regulation: Advances in Research and Theory Volume 4* (pp. 1–18). Springer US. https://doi.org/10.1007/978-1-4757-0629-1_1
- Rantanen, E. M., & Levinthal, B. R. (2005). Time-based modeling of human performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 49(12), 1200–1204. <https://doi.org/10.1177/154193120504901222>
- Rossi, V., & Pourtois, G. (2012). Transient state-dependent fluctuations in anxiety measured using STAI, POMS, PANAS or VAS: A comparative review. *Anxiety, Stress, & Coping*, 25(6), 603–645. <https://doi.org/10.1080/10615806.2011.582948>
- Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review*, 115(1), 101–130. <https://doi.org/10.1037/0033-295X.115.1.101>
- Smeding, A., Darnon, C., & Van Yperen, N. W. (2015). Why do high working memory individuals choke? An examination of choking under pressure effects in math from a self-improvement

- perspective. *Learning and Individual Differences*, 37, 176–182.
<https://doi.org/10.1016/j.lindif.2014.11.005>
- Soleimani Rad, M., Boroujeni, S. T., Jaberimoghaddam, A. A., & Shahbazi, M. (2022). Performance and decision making of a complex skill under monitoring and outcome pressure conditions: Which of them can reinvestment predict? *Psychology of Sport and Exercise*, 59, 102128.
<https://doi.org/10.1016/j.psychsport.2021.102128>
- Spielberger, C. D. (1970). Manual for the State-Trait Anxiety Inventory (self-evaluation questionnaire). (*No Title*). <https://cir.nii.ac.jp/crid/1370285712575158016>
- van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, 25(6), 2005–2015.
<https://doi.org/10.3758/s13423-018-1432-y>
- Wickens, C. (2021). Attention: Theory, principles, models and applications. *International Journal of Human–Computer Interaction*.
<https://www.tandfonline.com/doi/abs/10.1080/10447318.2021.1874741>
- Wickens, C. D. (1991). Processing resources and attention. In *Multiple Task Performance*. CRC Press.
- Wine, J. (1971). Test anxiety and direction of attention. *Psychological Bulletin*, 76(2), 92–104.
<https://doi.org/10.1037/h0031332>
- Wood, G., & Wilson, M. R. (2010). A moving goalkeeper distracts penalty takers and impairs shooting accuracy. *Journal of Sports Sciences*, 28(9), 937–946.
<https://doi.org/10.1080/02640414.2010.495995>
- Wright, C. A., Dobson, K. S., & Sears, C. R. (2014). Does a high working memory capacity attenuate the negative impact of trait anxiety on attentional control? Evidence from the antisaccade task. *Journal of Cognitive Psychology*, 26(4), 400–412.
<https://doi.org/10.1080/20445911.2014.901331>
- Yamanaka, K., & Kawakami, M. (2009). Convenient evaluation of mental stress with pupil diameter. *International Journal of Occupational Safety and Ergonomics: JOSE*, 15(4), 447–450.
<https://doi.org/10.1080/10803548.2009.11076824>