UiT The Arctic University of Norway

Department of Psychology, Faculty of Health Sciences

Exploring the Link between Mind Wandering and Reinforcement Learning through Behavioral Analysis and Pupillometry

Ceylin Yagmur Karayel

Master’s thesis in psychology (PSY-3900) - May 2024

Foreword

I remember the day I was accepted into the program, a moment of great joy for both me and my family. Since my arrival in Tromsø, I have been studying in the department of my dreams, enjoying every moment and keeping a sense of gratitude in my heart, even during the toughest times.

I have always been fascinated by learning cognitive neuroscience, where we explore the profound depths of our brains and learn the dynamics underlying our behaviors. This passion led me to express my interest in collaborating with Matthias Mittner and Gábor Csifcsák. They introduced me to a project involving topics that were initially unfamiliar to me. Despite all the challenges of understanding the complicated nature of reinforcement learning and mind wandering, I still wanted to be part of this project. Throughout this journey, Matthias and Gábor have been outstanding guides, always ready to assist and encourage me to further exploration. And most importantly, they taught me the importance of not only having ideas but also understanding the arguments behind them. I am deeply grateful to my supervisors for giving me the opportunity to work on this project together and for teaching me the true essence of 'learning' along the way, never withholding their support.

Thanks to my university, the wonderful people in my department, and Tove Dahl, who never lost her cheerfulness, as well as to my classmates. Most importantly, I want to thank my mother, who put in so much effort to bring me to this point of my life, my dear little brother, and my father who supported me every step of the way. Finally, I want to thank my boyfriend Besir, who has always been by my side on this journey. Your support and love are invaluable to me.

I will always look back on this journey with happiness and gratitude. I am proud of who I have become and what I have achieved over these last two years. Being a part of this project and living in Tromsø has been a truly enriching experience.

Ceylin Yagmur Karayel
Master student

Matthias Mittner
Supervisor

Gábor Csifcsák
Co-supervisor

Exploring the Link between Mind Wandering and Reinforcement Learning through Behavioral Analysis and Pupillometry

Ceylin Yagmur Karayel

PSY-3900

Department of Psychology – Faculty of Health Sciences

UiT, The Arctic University of Norway

Supervisor: Matthias Mittner

Co-supervisor: Gábor Csifcsák

Tromsø, May 2024

Sammendrag

Mange studier har vist de negative effektene av tankevandring på kognitiv ytelse, men effekten på forsterkende læring, en prosess basert på belønnings- og tapsbasert læring, er et uutforsket felt. I den nåværende studien hadde vi som mål å undersøke forholdet mellom tankevandring og forsterkende læring nærmere ved å bruke en ny forsterkende læringsoppgave for å observere hvordan tankevandring påvirker ytelsen. I tillegg brukte vi pupillometri for å få innsikt i nevrofysiologiske prosesser knyttet til stimulusbehandling, beslutningstaking og resultatevaluering. Vi antok at tankevandring ville være assosiert med suboptimale valg, som potensielt kunne påvirke motivasjon og læring. Resultatene våre viste at deltakerne var mer sannsynlig å oppleve tankevandring når den gjennomsnittlige belønningsraten var lav; dette støttet ideen om at tankevandring kan ha en motivasjonskomponent drevet av forventet belønning. Vi fant også at reaksjonstidene var raskere under tankevandring, noe som gjenspeiler en redusert tendens til å tenke før du tar et valg. I pupillometrianalysen observerte vi svekkede elevresponser under både stimulus- og belønningsbehandlingsfasene under tankevandring. Imidlertid fant vi større elevresponser i prosesseringen av både stimuli og belønning i oppgavetilstanden. Resultatene våre bekreftet at de negative effektene av tankevandring i mange kognitive domener også gjelder for forsterkende læringsoppgaver. Samlet sett ga undersøkelsen av forholdet mellom tankevandring og forsterkende læring, spesielt med tanke på at oppgaver med forsterkende læring iboende involverer belønninger av forskjellig størrelse, oss en unik mulighet til å vurdere forholdet mellom egen oppgaveverdi og endringer i oppmerksomhet.

Nøkkelord: tankevandring, forsterkende læring, pupillometri, motivasjon**.**

Abstract

Many studies have shown the negative effects of mind wandering (MW) on performance in sustained attention tasks, however, its effects on reinforcement learning (RL), a process based on reward- and loss-based learning, is an unexplored field. In the current study, we aimed to investigate the relationship between MW and RL by employing a novel RL task to observe how MW influences performance. In addition, we applied pupillometry to gain insight into how neurophysiological processes related to stimulus processing, decision-making, and outcome evaluation are related to MW. We hypothesized that MW would be associated with suboptimal choices, potentially by influencing motivation and learning. In line with our results, we found that suboptimal choices were associated with reduced performance, and the participants were more likely to experience MW when the average reward rate was low; this supported the idea that MW may have a motivational component driven by the expected reward. We also found that reaction times were faster during MW, reflecting a reduced tendency to think before making a choice. In the pupillometry analysis, we observed attenuated pupil responses during both the stimulus and reward processing phases during MW. Our results confirmed that the negative effects of MW in many cognitive domains also apply to RL tasks. Taken together, the investigation of the relationship between MW and RL, specifically considering that RL tasks inherently involve rewards of different magnitudes, provided us with a unique opportunity to assess the relationship between intrinsic task value and shifts in attention.

Keywords**:** mind wandering, reinforcement learning, pupillometry, motivation**.**

Exploring the Link Between Mind Wandering and Reinforcement Learning through Behavioral Analysis and Pupillometry

In the last century, the study of the sophisticated functions human brain has become one of the outstanding areas of cognitive sciences. This interest primarily arose from the rapid evolution of the human brain and its unique cognitive abilities that set humans apart from other primates (Vallender et al., 2008). Advances in technology and our increasing collective knowledge of brain imaging methods enhanced the popularity of the field and inspired generations of neuroscientists. Some research (Semendeferi et al., 2002) emphasizes the interconnections in the prefrontal cortex play a significant role in the development of the brain and specifically highlights the association of the prefrontal cortex with several higher cognitive functions such as decision-making, problem-solving, and attention.

One of the core components of human cognition is attention which is described as the focusing of our internal thoughts and conscious awareness (Cohen, 2014). This process allows individuals to selectively focus on one from various potential stimuli and filter out other irrelevant information concurrently (Cohen, 2014). However, sometimes our minds encounter challenges in focusing on a singular task, as both internal thoughts and feelings and external stimuli from the environment influence our attentional state of mind. Here in this study, we are examining the potential influence of MW, which refers to the state when our mind is disengaged from the current task, and whether this process has a motivational component driven by the expected reward in the decision-making task.

Theoretical Background

Mind wandering

In recent years, research focusing on the formation and components of spontaneous thoughts has been ever-increasing in the fields of neuroscience and cognitive psychology

(Bonifacci et al., 2023; Wong et al., 2023). A notable characteristic of the human mind is its tendency to generate thoughts when executive elements of attention are not occupied spontaneously (Antrobus et al., 1970). However, sometimes it can generate spontaneous thoughts even when it is occupied with cognitively demanding tasks (Thomson et al., 2014). The phenomenon denominated ‘mind wandering’ (MW) refers to when a person's attention and concentration drift away from the current task by being interrupted by inner thoughts such as retrospective and prospective thoughts, including personal concerns (Baird et al., 2011).

MW broadly comprises task-unrelated images and thoughts (TUITs) (Giambra, 1995), zone outs (Schooler, Reichle & Halpern, 2004; Schooler, 2002) and mind pops (Kvavilashvili & Mandler, 2004). To exemplify MW, we sometimes find ourselves generating internal thoughts completely detached from the environment and situation we are in. Various topics constantly arise in our minds; From deciding what to cook for dinner tonight, and thinking about future career goals, to reflecting on how great last summer vacation was. In these moments, our thoughts are entirely disconnected from the task at hand, spanning across the past, present, and future (Smallwood & O’Connor, 2011; Tulving, 1985; Baird et al., 2011). According to a study by Killingsworth and Gilbert (2010), individuals experience MW for nearly 40-50% of their waking hours in their daily activities such as reading, preparing food, or doing homework (Killingsworth & Gilbert, 2010).

Considering that both task-unrelated and task-related thoughts may rely on common cognitive resources (Smallwood & Schooler, 2006), the extent to which MW uses these resources depends on the cognitive demands of the task, such as task difficulty. Therefore, being preoccupied with internal thoughts can result in impairment in information processing and thus task performance (Esterman & Rothlein, 2019). It has been shown by many well-established studies that MW negatively impacts cognitive functions, especially executive functions, working memory, and attention. (Mooneyham & Schooler, 2013).

Many studies in the literature propose that MW tends to occur more in easy tasks compared to difficult tasks (Giambra, 1995; Seli et al., 2018; Smallwood et al., 2011; Thomson et al., 2013). In this respect, Smallwood & Schooler (2006) and McVay & Kane (2010) have developed two different perspectives as to the reason for the occurrence of MW. On the one hand, Smallwood et al. (2006) suggest MW is less likely to occur when the current task requires robust cognitive effort and attention, while people tend to engage in mind wandering when task demands are relatively low or automatic (Smallwood & Schooler, 2006). They also emphasize the reason behind the occurrence of MW in easy tasks stems from when task requirements are lower, thus other cognitive resources become available for generating MW (Smallwood & Schooler, 2006). On the other hand, when it comes to the 'why' of the occurrence of MW, McVay & Kane (2010) strongly argue that MW derives from the failure of executive control. This is because the executive control system fails to inhibit task-irrelevant thoughts or images that arise automatically (McVay & Kane, 2010).

Schad and colleagues (2012) suggest that the phenomenon of MW can be understood through two distinct phases of cognitive processing (Schooler et al., 2011). The first phase involves the drifting of attention away from the current surroundings, resulting in reduced processing of sensory information (Kam et al., 2011). This can lead to a decoupling of attention, ultimately impacting task performance (Christoff et al., 2009; Killingsworth & Gilbert, 2010). Secondly, MW is characterized by an internal state of mind, comprising task-unrelated thoughts and images, often arising from memory (Carriere et al., 2008; Smallwood & Schooler, 2006; Stawarczyk et al., 2011).

Although the cognitive costs of MW have been extensively studied in the literature (e.g., sustained attention (McVay & Kane, 2009), reading comprehension and eye movement (Reichle, Reineberg, & Schooler, 2010), and memory (Smallwood, Nind, & O’Connor, 2009), it has also been suggested to have benefits, which are connected with creativity, planning, and future-oriented thinking (Mooneyham & Schooler, 2013), and as well as improvements in statistical learning (Vékony et al., 2023). Baird, Smallwood, and Schooler (2011) conducted a study to investigate MW episodes during low-resource-demanding tasks and assess subjects' thoughts as either self-generated or goal-related. They found that subjects were prone to prospective thinking while their mind was wandering during the task. As a result, the study proposes that prospective MW may enable individuals to contemplate future-oriented goals and planning (Mooneyham & Schooler, 2013).

The scientific literature emphasizes the importance of the multidimensional nature of MW, as Giambra (1995) differentiated MW between two types: spontaneous (unintentionally) and deliberate (intentionally). Spontaneous MW refers to intrusions that spontaneously into your head without any effort, while deliberate MW occurs when individuals deliberately try to think about something other than the task at hand (Giambra, 1995; Seli et al., 2016). According to Seli et al. (2016), spontaneous MW is characterized by a lack of awareness, namely when people are unaware that their minds are wandering. In contrast, deliberate MW is associated with metacognitive awareness, and people are more likely to be conscious of their wandering thoughts (Seli et al., 2016), indicating that the individual is aware that their thoughts are diverting from the main task (Randall et al., 2014; Schooler, 2002).

In light of the research in the literature, both spontaneous MW and deliberate MW may lead to cognitive failure and impaired task performance (McVay & Kane., 2009; Schad et al., 2012), however, deliberate MW is also associated with non-task goals, such as creative and future-oriented thinking (Baird et al., 2012; Stawarczyk et al., 2011). Consequently, it is important to know the characteristics of the subdimensions of MW in different contexts, which can enhance our understanding of the implications of MW for cognitive functions.

Two noteworthy theoretical accounts have come to the forefront of the study of MW: the current concern hypothesis and the perceptual decoupling hypothesis. Klinger et al. (1973)

suggest that the reason behind the occurrence of MW is that people have preoccupations, desires, struggles, wishes, goals, etc. that go beyond their current task. According to the current concern hypothesis, when individuals perform a task, if they lack interest in the task and their current concerns are more valuable or rewarding than the task at hand, their minds naturally tend to wander toward self-generated thoughts (Randall et al., 2014). For example, when we are watching a movie that captures our interest, our attention is focused on this external event. However, if the movie is not exciting enough, our minds tend to generate internal thoughts that are more valuable than the movie. At this juncture, the magnitude of the value of our current concerns and the value of the current task interacts with each other and determine the degree of MW (Smallwood, 2013). In this respect, in such situations when a participant does not find the RL task interesting enough and thus has low motivation toward the task, we would expect the participant to start to think about their current concerns instead of focusing on the task.

Secondly, the perceptual decoupling hypothesis which is named by John Antrobus and Jerome Singer (Antrobus et al., 1970) involves the disengagement of attention or cognitive processes from external information and starting to focus on internal information, resulting in decrements in attention toward external information. This account describes the dynamics of MW, rather than the occurrence of the MW. Thus, we would anticipate reduced attention to external stimuli (i.e. RL task) during MW. Even though the association of MW with absent-mindedness, Schooler et al., (2012) propose MW is not the complete detachment from external information, on the contrary, it means being able to maintain the information from the external world in the same order as the internal information to ensure its continuity (e.g., reading).

As discussed above, we know that MW has various negative effects on cognitive functions. However, it has not been investigated in the literature so far whether MW also influences cognitive tasks that rely on reward or loss-related learning, such as RL. Therefore, in this study, we were interested in implementing a specific task to combine MW proximity in the RL in this study.

Reinforcement learning

Since the day we were born, we have been interacting with our surroundings. We interpret, observe, and regulate our behaviors within the contingency of cause and effect. From an evolutionary perspective, one of the most primitive learning methods is learning through feedback, known as reinforcement learning (RL). RL refers to a learning paradigm where we make decisions by evaluating potential outcomes that are associated with these decisions within a given situation (Rangel et al., 2008). We can easily relate this paradigm to our daily life experiences, for example, a person who touches fire for the first time realizes that it is a flammable substance and shows avoidance behavior the next time. All choices involve different consequences, and individuals proceed by exploring the different choices to achieve optimal results, even if this results in error. The context of RL involves two strategies of decision-making; 'exploitation' (Sutton & Barto, 2014) refers to a strategy when people take actions that have previously resulted in rewards; they are aware of the possible outcomes and expect a positive result. In contrast, the 'exploration' (Sutton & Barto, 2014) strategy of RL indicates when people try out new actions or choices in order to explore potentially better results that have not tried before. This concept brings with it a trade-off between exploitation and exploration because, in order to achieve optimal outcomes or rewards in a stochastic environment, it is necessary to make choices to ascertain their current value. Even though the exploitation strategy contains the actions that have been taken before and resulted in positive outcomes, individuals should try diverse actions to explore optimal ones.

The reward-based nature of RL creates a positive correlation between the value of the reward and one's motivation to pursue specific choices. As Schultz (2015) points out: "*Rewards are attractive. They are motivating and make us exert an effort. We want rewards; we do not usually remain neutral when we encounter them*."

Many studies in the literature have shown that reward-based learning has a positive effect on task performance in cognitive tasks. The study conducted by Padmala & Pessoa (2011) applied a selective attention task similar to the Stroop test. In this task, some trials began with a decision indicating a monetary reward for fast and correct answers, while other trials did not offer any reward. According to the results, participants performed better in reward trials in terms of both error rates and reaction times than in non-reward trials (Padmala & Pessoa, 2011). Therefore, we can assume that motivation is a significant factor that drives our behavior and actions to achieve rewards or positive outcomes. Hence, based on the research (Botvinick & Braver, 2015) on this subject strongly supports the notion that the reward-based learning system is linked to motivation and increased task focus. Since MW is characterized by a drift of attention from the current task, our hypothesis suggests that MW may affect the balance between 'exploitation' and 'exploration' in the RL task. That is, during task periods when reward rates are low, we would expect the likelihood of MW to increase. Therefore, we can establish a connection between the likelihood of experiencing MW and task engagement in situations where individuals perceive tasks as boring, as this indicates that joining in MW is perceived as higher than the perceived task value (Smallwood, 2013). In contrast, we would expect that during task periods when rewards are high, the likelihood of experiencing MW to be lower and, accordingly, behaviors will be more exploitative. In this case, we can hypothesize that when the value of the task increases, the probability of experiencing MW reduces. Considering both situations, we can argue that the relationship between the task value and the reward is related to our motivation and therefore to our likelihood of experiencing MW.

Until now, neurophysiological measurement methods have been widely used in both MW and RL research. There are many examples in the literature regarding the validity of MW and RL research, especially where neurophysiological measurement techniques are applied. In this regard, we investigated the interaction of MW and RL not only at the behavioral level but also included pupillometry in our study, which is a non-invasive technique. Particularly the close association between pupil size and various cognitive phenomena in the brain made us decide that pupillometry would be the most convenient measurement method for this study. Consequently, we examined the interaction of the MW and RL and utilized both behavioral measurement and pupillometry in this study.

Pupillometry and attention

Pupillometry has become a widely used method in disciplines such as psychology, neuroscience, and linguistics in recent years. The explanation for this is that the cognitive pupil responses provide us with a reliable estimate of the intensity of attention, perception, emotion, and other mental activities (Laeng et al., 2012).

Norepinephrine (NE) is a neurotransmitter responsible for brain functions and behavior, which is produced in the nucleus known as the Locus Coeruleus. The LC-NE system is primarily known for its effects on sensory processing and regulating arousal (Berridge & Waterhouse, 2003). Some research (Aston-Jones & Cohen, 2005; Devilbiss et al., 2006) emphasizes the LC-NE system can also be responsible for processes such as motivation and decision-making in the brain. The studies of the LC-NE system suggest that LC neurons fire in two different modes of activity: phasic and tonic (Aston-Jones et al., 1999). The phasic mode is associated with the activation of LC neurons in the presence of task-related stimuli

(Usher et al., 1999). In contrast, the tonic mode of LC is related to disengagement from the task at hand and provides a baseline level of activity (e.g., sleep, arousal), including distractive and exploratory behavior (Aston-Jones & Cohen, 2005; Laeng et al., 2012; Rajkowski et al., 1994). Notably, Aston-Jones & Cohen (2005) proposed that the phasic activity of the LC-NE system helps filter out irrelevant and distracting stimuli and facilitates the brain to focus on relevant signals by preventing irrelevant stimuli from impairing performance. Therefore, this argument indicates the phasic mode of the LC-NE system potentially enhances cognitive functions and optimizes task performance (Berridge & Waterhouse, 2003).

Furthermore, various studies (Verney et al., 2001; Joshi et al., 2016; Brocher & Graf, 2017) have demonstrated that the activity of LC-NE is closely related to pupil diameter. Aston-Jones & Cohen (2005) also suggest that stimulus processing is linked to rapid and dramatic pupil dilation, which is associated with the emergence of LC phasic response to stimulus-related events (Beatty, 1982). Given all the knowledge together, we would expect pupil dilation and more phasic changes in the LC-NE system during the RL task, in the moments where higher attention is present in the task (i.e., on-task condition). Therefore, the understanding of the connection between the phasic mode of the LC-NE system and pupillary responses enables us to measure changes in pupil dilation and capture event-related potentials (ERPD) during task performance.

The Current Study

Based on previous research, we know that MW has negative effects on cognitive functioning. However, we are still uncertain whether these negative effects also apply to RL. Thus, our main research question was:

"Does mind wandering negatively affect task performance in reinforcement learning tasks?"

To address this research question, the following six hypotheses were formulated:

Hypothesis 1 We expected self-reported MW scores (BMW-3) to be correlated with both the MW frequency and RL task performance.

Hypothesis 2 We expected the RL performance as reflected by choice optimality, to be lower when participants are engaging in the MW.

In addition to the choice optimality, RL performance was also evaluated by measuring reaction times (RTs). Therefore, we were also interested both in the average response speed corresponding to the participants' choices and in the magnitude of response speed variability.

We predict a higher RT coefficient of variability (RTCV) during RL and longer RTs in the reports of MW. Given the well-established association between the RTCV and MW (Bastian & Sackur, 2013), we anticipated that the RTCV was an indicator of a higher MW rate.

Hypothesis 3 We expected that participants would show longer RTs during MW as well as increased RTCV.

We further explored to see whether the average reward rate preceding each thought probe, which we believe to be related to the intrinsic value of the task for participants, will be predictive of their self-reported attentional states (on-task vs MW). To examine this further, we formulated a separate hypothesis:

Hypothesis 4 We expected a negative association between average reward and likelihood of self-report MW.

Hypothesis 5 We expected the participants would demonstrate attenuated pupil dilatations following stimulus onset while engaging in the MW. This would be associated with perceptual decoupling, as attenuated pupil dilations indicate compromised processing and evaluation of task-relevant environmental events.

Hypothesis 6 We expected the participants would demonstrate attenuated pupil dilations following reward onset while engaging in MW. This may be associated with a direction of mental resources to the MW state and reduced reward-related processing. This may cause reduced mental effort in calculating the optimal choice and pronounced processing of the reward.

Methods

Participants

The study sample originally consisted of 50 participants, 28 women and 22 men (average age range of women: 24.60, average age range of men: 26.77). Due to not choosing the MW state at all in the probe during the RL task, 10 participants were excluded from the t-test analyses, and the remaining 40 participants were evaluated. Likewise, 7 participants were excluded from the pupillometry analysis due to technical problems with eye data, and the remaining 43 participants were evaluated. All 50 participants participated in the questionnaire analysis, although,1 participant completed the experiment, the participant was removed from the experiment and replaced with a new participant due to recognition of the symbols in the task. Additionally, some participants did not complete the experiment because they did not meet the participant criteria on the day of the experiment. (e.g. too much eye makeup, incorrect vision, laser surgery, etc.). All the participants were recruited through a participant referral approach. This method involves leveraging word of mouth, including direct invitations to colleagues and classmates, as well as encouraging recruited participants to involve their acquaintances in the experiment. Participant consent was provided by all participants. The Internal Ethics Committee in the Department of Psychology, The Arctic

University of Norway approved the experiment. Participants were included in the experiment in cases they fulfilled following inclusion criteria; Participants should be between 18 and 60 years old, with unimpaired or corrected eyesight, and no history of psychological or neurological disorders (e.g., bipolar disorder, major depression, epilepsy, severe head injury, or brain surgery), no current medication that influences brain functioning (e.g., anxiolytics, antidepressants) and lastly, sufficiently rested before the testing day, have eaten enough and not be under the influence of any psychoactive drugs (e.g., alcohol, narcotics). One of the participants was excluded from the experiment due to recognizing the meaning of the symbols in the experiment and was replaced with a new participant.

Materials and procedure

Questionnaires

Need for Cognition (NFC) questionnaire

The NFC was conceptualized by Cacioppo and Petty (1982). The most widely used and recognized version of the Need for Cognition scale consists of 18 items. The NFC assesses the tendency of people to engage and enjoy events that require cognitive effort (Leary & Hoyle, 2013). Sadowski and Gulgoz (1992) measured internal consistency and test-retest reliability of the Need for Cognition Scale, and they found the internal consistency of Cronbach alpha was 0.91 in the first administration and 0.92 in the second administration of the questionnaire. They also found that the questionnaire demonstrated consistent and high test-retest reliability over the 7-week time span. ($r = 0.88$, $p < 0.0001$). Participants have chosen their level of agreement or disagreement with the statements using a 5-point scale in Likert format with 1 indicating ‘Very strongly disagree’ and 5 indicating ‘Very strongly agree’.

The Brief Mind Wandering Three-Factor Scale (BMW-3)

BMW-3 is one of the most recent scales for measuring different aspects of MW which was designed by Schubert et al. (2023). The BMW-3 defines MW as thoughts unrelated to the task at hand and assesses three dimensions of MW: Unintentional MW (UI-MW, e.g., “When watching TV, other things inadvertently cross my mind”), intentional MW (I-MW, e.g., “I deliberately allow my mind to wander to escape the daily grind”), and meta-awareness of MW (MA-MW, e.g., “I immediately notice when my thoughts are not in the here and now”). In the experiment, participants are informed they will encounter a questionnaire that consists of 12 statements that describe typical situations related to individuals’ thoughts in daily life. Participants are instructed to thoroughly read each statement and express their level of agreement or disagreement. The BMW-3 items are evaluated using a 5-point Likert scale, where 0 corresponds to 'Fully Disagree,' and 4 corresponds to 'Fully Agree’. Moreover, they are told that there are no right or wrong answers, and they should just answer intuitively. The internal consistency of the three scales were found to be within acceptable to good ranges in the English version of the scale. Specifically, $\alpha = 0.72$ for the UI-MW scale, $\alpha = 0.83$ for the I-MW scale, and $\alpha = 0.79$ for the MA-MW scale.

Participants completed the questionnaires with a website tool Nettskjema.no which is a platform designed for creating and conducting online surveys. All the data were securely stored in an anonymized manner, identified only by participant numbers.

Study setup

Following the questionnaires phase, all participants entered in a quiet, separate, and well-lightened chamber, which belongs to the eye-tracking lab at the Department of Psychology, UiT The Arctic University of Norway. Our Open Science Framework (OSF) repository contains all materials and anonymized data for public access https://osf.io/ca95r/ . In accordance with the experimental procedure, none of the participants had any electronic devices or anything else that would cause distraction or affect the flow of the experiment.

Experimental design

Participants performed a novel RL task, that was adapted from Frank et al., (2004). The task was designed to measure the impact of changing stimulus values on task performance and the frequency of MW. In the experiment, participants were presented with pairs of unfamiliar symbols, each assigned a numerical value. The set of all presented symbols consisted of six different Chinese characters, each of which was assigned a numerical value ranging from 0 to 20. The main goal of the task was choosing the symbols that had the highest value by following a trial-and-error strategy. The pair of symbols were randomly positioned on the left and right sides aligned on the screen. The symbols, each covering 2/10 of the screen width, were positioned at a visual angle of approximately 5.65 degrees apart from each other when viewed from a distance of 70 cm.

Throughout the task, all possible pairs of symbols taken from the set of six symbols were presented over 300 trials in total (see Figure 1). In each trial, a single pair was presented in a randomized order. The values of the symbols were completely independent and random, and each symbol was randomly positioned either to the right or left in each trial. The value of the symbols was not known beforehand, and the numerical value of a symbol was disclosed upon selection, while the value of the unchosen symbol remained hidden.

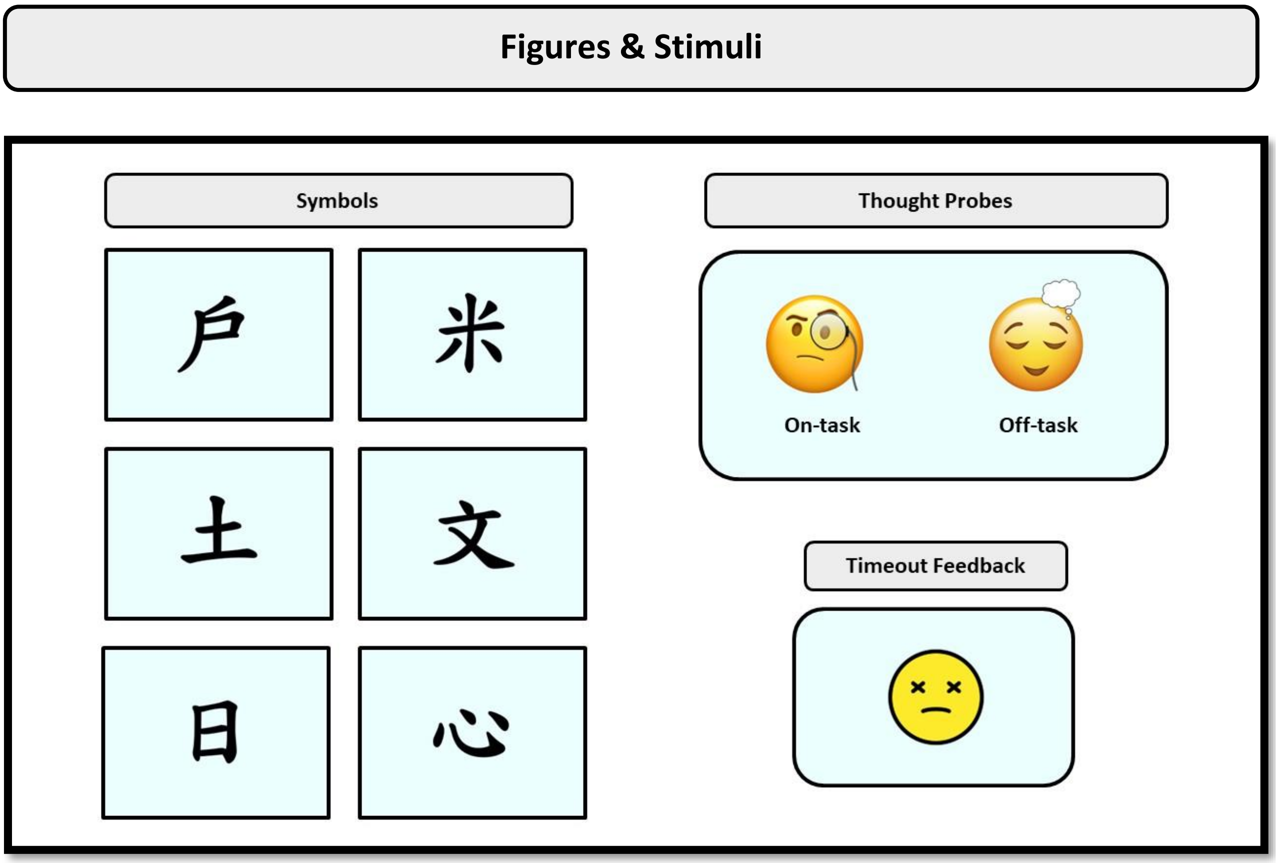


Figure 1. Figures and stimuli of main task

The values of the symbols indicated how many points were earned from this symbol (see Figure 2). The value of the symbols was designed to gradually change throughout the task following a random walk. This means, that the current value of each symbol changed by a random value taken from a zero-mean normal distribution (SD = 2) at each trial. The program would generate a random integer with a standard deviation of 2 that was centered at 0. This integer would be added to the symbol’s existing value (i.e. 3 would add to a previous value of 8 to make 11, while -2 would add to a previous value of 8 to make 6). If this calculation would lead to the value of the symbols to exceed the boundaries, which were set at 0 and 20, the remaining added value would be reflected in the opposite direction (i.e. 3 would add to a previous value of 19 to make 18, while -2 would add to a previous value of 0 to make 2).

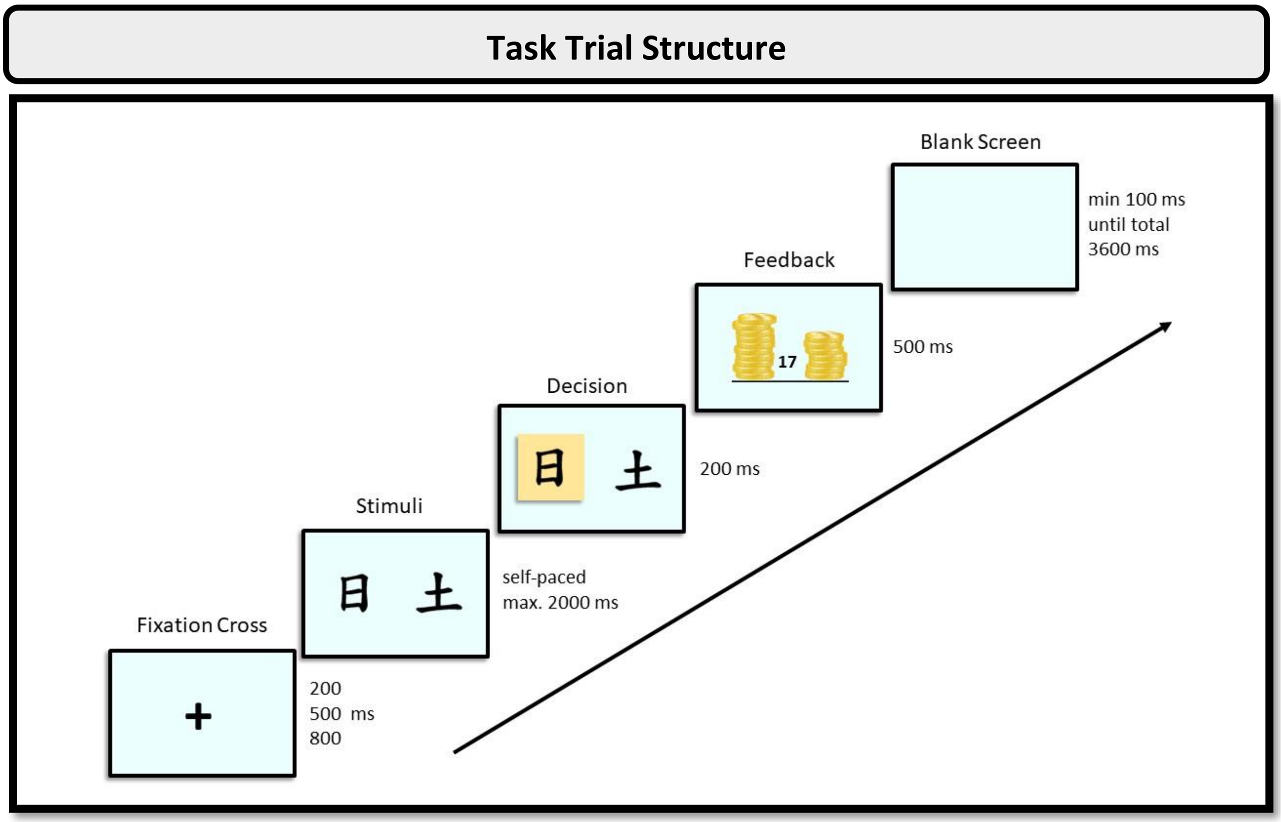


Figure 2. Visualized example trial of main task

Each trial started with presenting a fixation cross positioned in the middle of the screen with a duration randomly chosen to be either 200, 500, or 800 milliseconds. Following this, the participants were shown the pair of symbols and requested to choose one of the symbols within 2000 milliseconds. The participants were instructed to choose the symbols by pressing the 'F' key for the left symbol and the 'J' key for the right symbol with their index fingers. The chosen symbol was highlighted for 200 milliseconds. However, if the participant had not chosen a symbol within 2000 milliseconds the trial was discontinued displaying an emoticon representing a 'TIMEOUT' for 200 milliseconds. If the participant selected a symbol within a determined time, they were shown the numerical feedback of the symbol together with a stack of coins representing the number of points for 500 milliseconds. Before moving to the next trial, a blank screen was displayed lasting a minimum of 100 milliseconds and persisting until the trial duration of 3600 milliseconds was achieved (see Figure 2). The background in the original task was gray, and all images were colored in dark red. The color red was chosen for our experiment because the adaptation/recovery of the pupil due to light exposure is faster for red compared to other light (see Mathot, 2018).

In this experiment, we implemented thought probes as a self-report of MW throughout the task trials. Participants assessed their current thoughts by choosing expressions consisting of a pair of emoticons. While the emoticon positioned on the left represented task- focus (on-task), the emoticon on the right represented MW (off-task) (see Figure 1). During the task, the participants were repeatedly interrupted by thought probes and requested to select one of them as an indicator of their current attentional state. After selecting one of the emoticons presenting their current attention state, participants were shown a thumbs-up image as feedback for either decision. Throughout the 300-trial task, the thought probes were displayed at an average rate of one thought probe per 10 trials. The variability in their presentation was arranged ±5 trials, meaning they could appear randomly every 5 to 15 trials. As a result, the total number of trials with thought probes was constant at 30 probes during the course of the entire experiment. Before the main task, all the participants were requested to perform the training task as a familiarization of the main task. The training session involved the same procedure as the main task, including 5 thought probes (with a variation of ±1). The entire session lasted approximately 1 hour, encompassing the questionnaires (5-10 minutes), eye tracking calibration (5-10 minutes), task instructions, training session, and the main task (18 minutes).

Eye-tracker configuration and setup

For pupillometric measurements, we utilized a versatile desktop-mounted video-based eye tracker (EyeLink 1000 Plus, SR Research) that can simultaneously record both pupil sizes at up to 2000 Hz during the trial presentation. The eye tracker was connected to a desktop running EyeLink 1000 plus software on PsychoPy (Peirce, 2007), and the task was created and programmed using PsychoPy**.** The setup consisted of two main computers; (a) the Host PC which is responsible for executing real-time eye-tracking at 500 samples per second while computing the true gaze position on the display (Eyelink Plus 1000 Handbook, 2005-2009). At the same time, (b) the Display PC is tasked with presenting the stimuli (e.g., the reinforcement learning task) throughout the experiment (see Figure 3) and storing the behavioral data.

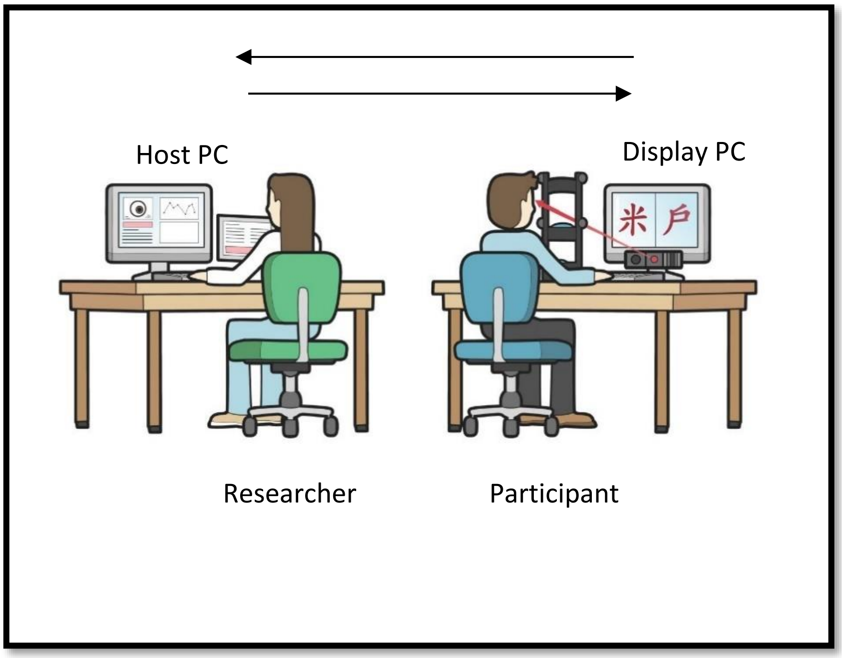


Figure 3. Illustration of the experiment setup

Participants sat on a chair and placed their chin on the chinrest in front of the eye tracker, which is connected to the monitor displaying the task. The task was presented on a 17-inch screen using an eye tracker, positioned at a distance of 60mm-70mm from the participants. The seat and chinrest were specially adjusted to ensure that each participant had the most comfortable position and head stabilization. In addition, the view of the eye-tracking camera was optimized to detect the pupil and corneal reflection (CR), ensuring the acquisition of clear and crisp images from each participant, and thus detection threshold or biases established. The pupil threshold levels must be between min 75 and a maximum of 115, and the corneal reflection (CR) levels must be a maximum of 240.

After these modifications, the software would initiate the calibration process. Calibration is used to gather fixation samples from predetermined target points to establish a mapping between raw eye data and gaze position. Participants are instructed that start the calibration by pressing a spacebar key. The Display PC presents target points (e.g., dots) sequentially, starting from the middle of the screen and then moving to the edges of the screen. Participants follow each dot in order and feedback graphics are then displayed on the Host PC. In this tracking mode, the calibration type consisted of 9 points. The calibration is automatically checked by the Host PC when finished and diagnostic information is delivered. The calibration process must be performed after camera setup and before the validation process. Following the calibration phase, the software would perform a validation process to ensure calibration accuracy.

Upon completing the calibration phase, the software performed a validation process to verify the accuracy of the calibration. Participants were instructed that start the calibration by pressing a spacebar key. The validation process followed the same procedure as calibration with a difference of refixation of two target points with the greatest error margins. The validation was automatically checked by the Host PC when finished with reporting the average and maximum error value for each eye which was predefined as below and equal to $\leq$0.5 for the average error and below to 1.0 for the maximum error. When the validation process succeeds, the margin range was highlighted in green, signifying 'good', however, if the validation error was large, the margin range was highlighted in red, indicating 'poor.' In cases where the validation fell within acceptable limits, the margin range was displayed in grey, representing 'fair.' If the accuracy at a fixated position was not acceptable, the calibration process was performed again by readjusting the threshold values.

Statistical Analysis

Between-subject Analysis

For testing the hypotheses regarding a relationship between RL task measures and our questionnaire data (hypotheses 1 and 2), we calculated Pearson correlations for variables of the BMW-3 subscores (i.e., unintentional MW, intentional MW, and meta-awareness MW), the NFC score, MW rate, and percentage points in the RL task.

The variables included in the correlation analysis were calculated individually across all trials as follows: "MWrate" indicates the percentage of MW during RL and was calculated as the number of MW probes divided by the total number of probes and multiplying the result by 100%. "perc_points" (percentage points) were calculated by subtracting the minimally achievable points from achieved points and then dividing the result by the difference between the maximum number of points and the minimally achievable points. All the behavioral analyses were calculated using the R programming language (R Core Team, 2023).

We calculated descriptive and inferential statistics for variables calculated on the probe-level. The "Loss" (i.e., choice optimality in the RL task) variable which indicates the average loss of possible points across the last 5 trials before each probe, was calculated by subtracting actually achieved reward from the best possible reward. We also calculated the average reward "Avgrew" variable before each probe as the mean reward during the last 5 trials before each probe. Finally, the "MeanRT" variable refers to the mean reaction time for the last 5 trials before each probe.

For each of these variables quantified the mean differences between MW and on-task conditions (i.e., when participants focus on the RL task). We included these variables in logistic regression models where we predicted the likelihood of MW using these variables as predictor variables.

Within-subject Analysis

In order to test the hypotheses regarding the task analysis, we utilized a random effects logistic regression analysis with a random intercept at participant-level, with responses to thought probes (coded as 0 = on-task, 1 = MW) as an outcome variable. In addressing hypothesis 2, we included two predictors in the model: Trial (Probe-number) and Loss variables. "Trial" variable refers to the trial number when each probe was presented and was adjusted using Z-scores, indicating that 0 is the middle of the experiment (trial 150), and one unit reflects one standard deviation (SD) before or after this midpoint. Each SD corresponds to approximately 87 trials.

For hypothesis 3, the model included two variables: the MeanRT and RTCV variables. The "RTCV" variable represents the reaction time variability for the last 5 trials before each probe, which was calculated by the standard deviation of the reaction time divided by mean reaction time (SD of RT / Mean RT). Besides, we anticipated the RTCV to be more variable and the mean reaction time to be longer during the RL task. Lastly, for hypothesis 4, we intended to consider the Avgrew variable separately, expecting there would be a negative association between the Avgrew variable and MW in the RL task.

Pupillometric Measurements and Preprocessing

The analysis and preprocessing of the pupillometric data were implemented using a Python package called ‘Pypillometry’ (Mittner, 2020). The unprocessed pupillometric data from both eyes were included in the analysis and then the eye with the fewest number of missing data points was chosen for each participant. In the pupillometric data, sometimes missing data occurs when the eye tracker fails to detect pupil size (Mathôt & Vilotijević, 2022), usually because the participant moves their head from their chinrest, closes their eyes completely, etc. In the case of blinking, the eye tracker records the pupil, but since the measured pupil size does not correspond to the actual pupil size, it is identified as invalid data (Mathôt & Vilotijević, 2022). To address this issue, we applied an eyeblink detection procedure developed by Mathôt (2013) that reconstructs eyeblinks so that they are not considered invalid. The aim of this algorithm is to reconstruct the pupil size during blinking while detecting and eliminating the non-blink artifacts. Next, the algorithm first determines the start and end points of each blink, and then reconstructs the missing signal. The reconstructed data marked as missing data in the signal for further analysis (Mathôt, 2013). In our pupillometric analysis, the pupil data was reconstructed while also minimizing the non-blink artifacts and was optimized for each individual participant. Parameters were tuned according to the individual characteristics of each participant. In this procedure, the raw and preprocessed signals were visually analyzed and evaluated by an expert analyst (MM). Then, blinks emerging within close together in time (<100 ms) were combined to avoid interpolation artifacts. After, blinks were linearly interpolated using the method described in Mathôt (2013), a 5 Hz lowpass filter (Butterworth) was applied to the continuous data. In addition, we excluded 7 datasets due to too many blinks and/or bad quality from the pupillometric analyses.

For our main analysis, we calculated the event-related pupil dilation (ERPD) based on the onset of the stimuli. We selected all segments of pupillometric data from -500 ms before stimulus onset until 3600 ms post stimulus and performed baseline-correction in the time-window [-200, 0] ms. For the general analysis of the stimulus-induced ERPD, we averaged all segments within-participant and then calculated the grand mean across participants (Figure 5 upper plot). To analyze the ERPD difference between on-task (OT) and MW, we calculated within-participant ERPDs separately for MW and OT using the 5 trials before each probe. These per-subject and per-condition ERPD curves were averaged across participants to produce the grand-mean curves per condition displayed in Figure 5 (lower plot). For significance testing, we calculated the difference curve on the individual subject-level and conducted pointwise one-sample t-tests. The figure highlights regions where this test was significant at the alpha = 0.05 level.

Results

We first looked at the participants' average rate of MW during the RL task (see Figure 4). Based on results from the behavioral data, the average percentage of MW was 25.4%, and the Standard Deviation (SD) was 24.5%.

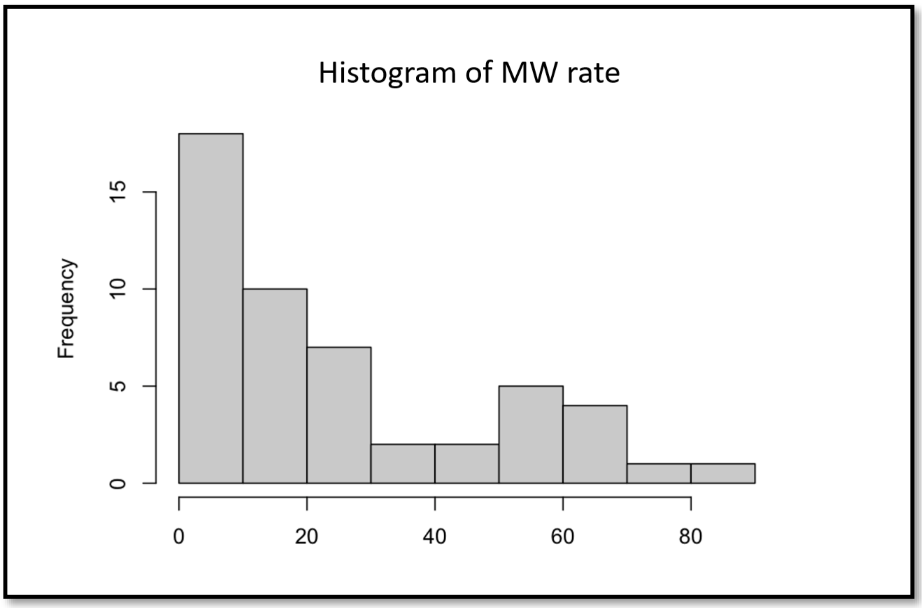


Figure 4. Histogram of MW rate during the RL task

The distribution was highly skewed (see figure 4), and the median was 16.7%, the skewness is 0.78. There were 10 participants who never reported have been MW during the task. Afterward, we looked at questionnaire analysis to examine the relationships between the variables we were interested in in our study.

Between-subject Analysis

Table 1 demonstrates a correlation to assess the association between self-reported questionnaire data and variables extracted from the task (N=50).

Table 1

Means, standard deviations, and correlations with confidence intervals

Variable	*M*	*SD*	1	2	3	4	5
1. BMWu	10.14	3.11					
2. BMWi	9.36	3.86	0.32* [0.05, 0.55]				
3. BMWm	7.54	1.50	-0.13 [-0.39, 0.15]	0.26 [-0.02, 0.50]			
4. NFCsum	63.78	10.10	-0.05 [-0.33, 0.23]	0.04 [-0.25, 0.31]	0.04 [-0.24, 0.32]		
5. MWrate	25.40	24.46	0.31* [0.04, 0.54]	0.16 [-0.12, 0.42]	-0.26 [-0.50, 0.02]	-0.27 [-0.51, 0.01]	
6. perc_points	66.23	7.58	-0.02 [-0.30, 0.26]	0.00 [-0.28, 0.28]	0.05 [-0.23, 0.33]	0.35* [0.08, 0.58]	-0.10 [-0.37, 0.18]

Note. *N* = 50. *BMWu* = Unintentional MW, *BMWi* = Intentional MW, *BMWm* = Meta-awareness MW, *NFCsum* = scores of NFC scale, *MWrate* = Average percentage of MW, *perc_points* = Percentage points. *M* and *SD* are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). * indicates $p < .05$. ** indicates $p < .01$.

A significant positive correlation was observed between unintentional MW (BMWu) and intentional MW (BMWi), $r = 0.32$, $p = 0.023$, 95% CI = [0.05; 0.55], however, there was no significant relationship between meta-awareness of MW (BMWm) and unintentional MW, $r = -0.13$, $p = 0.368$, 95% CI = [-0.39; 0.15]. Similarly, the result did not reach statistically significance between meta-awareness MW (BMWm) and intentional MW, r = 0.26, p = 0.071, 95% CI = [-0.02; 0.50]. For the correlation analysis of NFC, there was no significant association between NFC and any of the MW subscales: NFC and unintentional MW, r = -0.05, p =0.712, 95% CI = [-0.33; 0.23], NFC and intentional MW, r = 0.04, p = 0.806, 95% CI = -0.25; 0.31] and NFC and meta-awareness MW, r = 0.04, p = 0.760, 95% CI = [-0.24; 0.32]. With respect to the associations between questionnaire data and MW reports during the RL task, a significant positive correlation was observed between MW rate and unintentional MW, r = 0.31, p = 0.028, 95% CI = [0.04; 0.54], however, there was no statistically significant correlations between MW rate and intentional MW, r = 0.16, p = 0.270, 95% CI = [-0.12; 0.42]. Notably, there was a moderate negative correlation between the MW rate and meta-awareness MW r = -0.26, p = 0.0714, 95% CI = [-0.50; 0.02], as well as there was a moderate negative correlation between MW rate and NFC scores, r = -0.27, p = 0.061, 95% CI = [-0.51; 0.01] even though neither of these effects reached significance. A significant correlation was observed between percentage points and NFC scores, r = 0.35, p = 0.012, 95% CI = [0.08; 0.58]. However, there was no statistically significant correlation between percentage points and unintentional MW, r = -0.02, p = 0.878, 95% CI = [-0.30; 0.26], and similarly with the intentional MW, r = 0.00, p = 0.999, 95% CI = [-0.28; 0.28] and, no significant correlation between percentage points and meta-awareness MW, r = 0.05, p = 0.714, 95% CI = [-0.23; 0.33]. Finally, no significant correlation was found between percentage points and MW rate during the task, r = -0.10, p = 0.487, 95% CI = [-0.37; 0 .18].

Table 2 shows descriptive statistics for the variables calculated for each thought probe response during the RL task (N = 40). As can be seen from the table the mean value of the Loss (Choice optimality) variable was significantly higher (t (39) = 3.28, p < 0.01) for the MW condition (M = 7.10, SD = 1.48) than for the on-task condition (M = 6.32, SD = 0.77).

However, the mean value of the Avgrew (Average reward) variable was significantly lower (t (39) = 2.11, p < 0.05) for the MW condition (M=10.6, SD=1.57) than on-task condition (M = 11.2, SD = 1.11). Lastly, the mean value of the MeanRT (Mean reaction time) variable was significantly lower (t (39) = 2.29, p < 0.05) for the MW condition (M = 0.90, SD = 0.19) than for the on-task condition (M = 0.93, SD = 0.15).

Table 2

Descriptive Statistics

	MW		On-task						
Variable	***M***	***SD***	***M***	***SD***	***t***	***df***	***p-value***	***%95 CI***	***SE***
Loss	7.10	1.48	6.32	0.77	-3.2831	39	0.002	[-1.23, -0.29]	-0.76
Avgrew	10.6	1.57	11.2	1.11	2.1173	39	0.041	[0.02, 1.08]	0.55
MeanRT	0.90	0.19	0.93	0.15	2.2936	39	0.027	[0.004,0.06]	0.035

Note. N = 40, *M* = mean, *SD* = standard deviation, *SE* = standard error of the mean, * indicates *p* < .05. ** indicates, *p* < .01, *** indicates *p* < 0.001

Here we present a joint table (Table 3) providing a comprehensive summary of our four logistic regression models. This table encapsulates all model coefficients alongside overarching 'fit' measures at the bottom. Each model was represented within the first row of the table, with the variables we observed within the logistic regression models aligned along the first column in the table. To test our within-subject hypotheses, we applied four logistic regression models, treating MW as the dependent variable. Logistic regression used to test the probability of an event occurring depending on independent variables. This analysis method is preferred in situations where the dependent variable is 'binary', such as “0 or 1”, or “yes or no". In our RL task, we have such a condition, as only 2 response options when measuring MW, 'on-task' or 'off-task.

Table 3

Joint table for four logistic regression model

	(1)	(2)	(3)	(4)
(Intercept)	-2.57***	-0.78	-0.97*	-2.17***
	(0.34)	(0.53)	(0.39)	(0.55)
scale(trial)	0.63***	0.57***	0.62***	0.63***
	(0.08)	(0.08)	(0.08)	(0.08)
loss	0.12***			0.10**
	(0.03)			(0.03)
rtcv		0.33		
		(0.78)		
meanrt		-1.19*		
		(0.46)		
avgrew			-0.08**	-0.03
			(0.02)	(0.03)
SD (Intercept subj)	1.87	1.85	1.86	1.87
R2 Marg.	0.069	0.062	0.061	0.069
R2 Cond.	0.548	0.540	0.542	0.549
AIC	1283.1	1296.7	1292.3	1284.2
BIC	1304.3	1323.2	1313.6	1310.8
ICC	0.5	0.5	0.5	0.5
RMSE	0.34	0.35	0.35	0.34

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The first model (1) included the Trial number and the average loss as predictor variables in the model. According to the results of the first model, a positive coefficient value of the Trial coefficient ($b = 0.62$, $SE = 0.07$, $z = 8.20$, $p < 0.001$) indicates that MW was more likely late in the task. This refers to the time-on-task effect which is a well-known and robust finding in MW research (Thomson et al., 2014). A positive coefficient value of the Loss ($b = 0.12$, $SE = 0.02$, $z = 4.28$, $p < 0.001$) indicates that each unit increase in the loss variable is related to an increase in the probability of MW. Higher loss values due to choosing lower value symbols indicates worse RL performance, which means suboptimal choices. Therefore, this positive relationship showed that participants' task performance was lower before reporting MW. These results are consistent with our hypothesis 2.

In our second model (2), we investigated whether more impulsive choices associated with decreased RTs were related to shifts in attentional state. At the same time, we also wanted to test the hypothesis that MW would be related to higher behavioral variability, indexed by the RTCV variable. In contrast to our hypothesis, the estimated coefficient for mean reaction time was statistically significant ($b = -1.188$, $SE = 0.46$, $z = -2.56$, $p < 0.01$), which was opposite to the expected effect in the hypothesis. This negative relationship indicates that when the participants have faster reaction times, they are more likely to experience MW. A positive coefficient value of the Trial coefficient ($b = 0.57$, $SE = 0.07$, $z = 7.40$, $p < 0.001$) was found significant as an indicator of time-on-task effect. On the other hand, interestingly, we could not find a significant association between the RTCV and MW, as a predictor of MW ($b = 0.33$, $SE = 0.78$, $z = 0.42$, $p > 0.6$).

In our third model (3), we tested our hypothesis 4 to examine whether a negative association between MW and average reward was present. The results of the logistic regression analysis demonstrate a positive coefficient value of the Trial coefficient ($b = 0.61$, $SE = 0.07$, $z = 8.11$, $p < 0.001$) was found significant as an indicator of the time-on-task effect. The result for the average reward variable shows that a higher average reward was associated with a lower probability of MW ($b = -0.07$, $SE = 0.02$, $z = -3.02$, $p < 0.001$).

Accordingly, this result shows that in cases where the average reward rate of the last 5 trials preceding the probe was high indicates that participants report being on-task rather than engaging in MW. This result aligns with the context of hypothesis 4 - we can suggest that MW was more likely to occur in low-average reward situations.

Finally, we included an exploratory model (4) in our results. This model included both average reward and loss variable in order to disentangle their relative on MW. The reasoning for this analysis is based on the fact that our "loss" variable includes aspects of both task-performance (choosing the suboptimal stimulus) and the current (random) reward rate of the symbols. Therefore, we wanted to evaluate whether the loss variable would still be related to MW when controlling for the effect of the reward rate (by including average reward as a predictor). According to the results, the estimated coefficient value of the Trial coefficient ($b = 0.62$, $SE = 0.07$, $z = 8.11$, $p < 0.001$) was found significant as an indicator of the time-on-task effect. However, the estimated coefficient for the Avgrew variable was not significant ($b = -0.02$, $SE = 0.02$, $z = -0.92$, $p > 0.05$), while we found a significant effect for the Loss variable ($b = 0.10$, $SE = 0.03$, $z = 3.19$, $p < 0.001$). These findings show that even when the average reward variable is controlled, loss is still a significant variable.

Pupillometry Analysis

In Figure 5, we display the average event-related pupil dilation in response to the presentation of the stimuli at time zero over time. The graph represents both the processing of symbol pairs that participants had to choose from and the processing of reward that participants received after the choice was made, subsequently. The upper plot represents the average pupillary response of 300 trials per participant, and the lower plot is based only on data preceding the 30 probes.

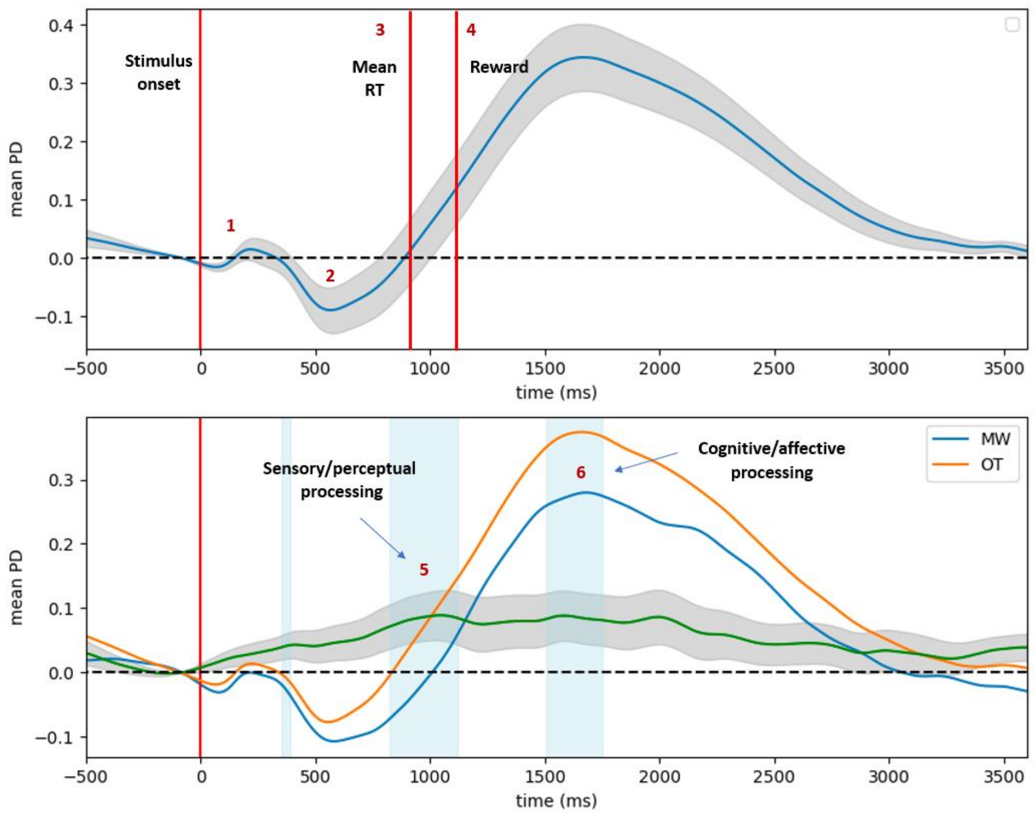


Figure 5. The average pupillary responses for per trial

Note. Blue curve indicates mind wandering condition, orange curve indicates on-task condition, green curve indicates difference between the curves. The shaded area depicts the standard error of the mean (SE). 1 = Artifacts (due to eye-movements), 2 = Light reflex, 3 = Mean RT, 4 = Reward, 5 = Sensory processing, 6 = Cognitive processing

We can identify and label the different components of this graph. The first inflection (1) around 50-200ms, we interpret as artifacts stemming from the prominent eye-movements away from the central fixation cross that were necessary to perceive the two stimuli. When the eye is turned away from the camera, the projection of the pupil on the camera plane is reduced due to an effect known as "foreshortening" (Hayes & Petrov, 2016), thereby producing an artifactual reduction in pupil size. The second major through (2) around 600ms, we interpret as pupil light reflex (Mathôt & Vilotijević, 2022) due to the lighting changes caused by the display of the stimuli. When the eyes are triggered by the light it takes time for the pupil to react to changes in light. After the pupil light reflex, (3) the average response time of the participant in choosing the symbols corresponds to around 900ms for MW and 930ms for OT condition, which means this process was most probably reflecting the decision-making process, indicating the moment when stimuli reached the pupil and the processing of the stimuli. (4) The following red line demonstrates when the reward appears on the screen, which corresponds to approximately 1100-1200ms in the trial, and then the processing of the reward is represented as cognitive and affective processing. In Figure 5 (lower), we show the ERPD following stimulus presentation separately for trials before probes that were answered as MW or on-task. The blue curve represents the mind wandering (MW), the orange curve indicates the on-task (OT) condition, and the green curve is the difference between the MW and OT curves (OT-MW). (5) Blue-shaded areas indicate significant time windows in which pupil dilation patterns differ from each other for MW and OT conditions. These results indicate that we found significantly attenuated pupil responses to the symbols in the MW condition, while significantly stronger pupil responses were observed in the OT condition (5). Specifically, the ERPD peak has a shorter onset and larger magnitude for OT relative to MW. (6) As can be seen in the second blue-shaded area, we observed a significant pupil response and dilatation for the OT condition during the cognitive and affective processing of the reward. Compared to the OT condition, we observed attenuated pupil responses during the processing of the reward in the MW condition.

Discussion

Behavioral Analysis

The main aim of this study was to investigate whether mind wandering negatively affects RL task performance. As hypothesized, we found significant effects that confirm the negative influence of MW both on behavioral measures and neurophysiological processes underlying of RL, linked to task performance and pupillary responses, respectively.

All the participants completed the BMW-3 questionnaire before proceeding to the main experiment, which measures their propensity towards MW in everyday life. The significant correlation between unintentional and intentional MW (BMWu and BMWi subscores respectively) confirmed the results obtained for the validity of the subscales by Schubert et al. (2023). However, unlike Schubert et al. (2023), we did not find a significant correlation between the meta-awareness dimension (BMWm) and other subscales in our study. Nonetheless, Schubert et al. (2023) emphasized that the correlation between BMWm and other subscales was lower strength compared to the correlation between BMWu and BMWi, and suggested that the reason for this was the fact that the meta-awareness might rely on broader aspects of cognitive processes (Schubert et al., 2023).

Regarding our hypothesis 1, we found a significant correlation between the frequency of reporting MW in thought-probes during the RL task and unintentional MW in the BMW-3, but not with intentional MW. While the positive association between the everyday propensity to unintentionally engage in MW and MW frequency in the lab during our task was anticipated, the absence of such a relationship for the intentional MW subscore was rather unexpected. To exemplify it, we can assume that participants would be motivated to perform well in an experiment conducted in a well-controlled experimental environment similar to ours, where the task is driven by motivation and requires cognitive engagement and alertness about the task (Liu et al., 2023; Seli et al., 2015). This might be because the task demands are cognitively high and require constant attention during the task. Hence, the absence of a significant correlation between intentional MW and MW rate during the task may be explained by this putative lack of intentional MW during our task. In addition, according to a study conducted by Robinson & Unsworth (2018), the reduced alertness was primarily due to unintentional MW, and intentional MW was primarily due to a lack of motivation in the study. These results are consistent with our argument and give us the answer to why we did not expect intentional MW in the RL task.

We can make different assumptions in line with this result; because the RL task we designed involves a reward-based learning approach, it includes rewards that are constantly presented and emotionally salient to the participants. This shows that the participants had an internal motivation for the task, which may have prevented engaging in MW episodes deliberately. The reason for this could be that the RL task had a faster pace than typical MW tasks (e.g., SART), and a forced-choice was required at each trial (as compared to the monotonous Go/NoGo nature of the SART, which highly relies on automatic responses). That is, in contrast to other tasks used in conjunction with MW probes, the RL task was probably more engaging. Therefore, when combined with the potential effect of rewards motivating participants, it may have led to the absence of the correlation between MW reports and the predisposition for intentional MW. To gather all these reasons together, we can suggest that the frequency of MW in the RL task why was not related to intentional MW.

According to our hypothesis 2, we found a significant relationship between the choice optimality and the outcome measure. Accordingly, when the participants' task performance was more varied, it meant they were less likely to choose the optimal option. The reason for this might be that the changing values of the stimulus during the task are not known to the participants and therefore, participants need to actually explore seemingly suboptimal stimuli to check how their values change over time. Especially in the RL task, which has a probabilistic structure, it is necessary to find a balance between 'exploitation' and 'exploration' behaviors to achieve optimal performance. As a matter of fact, disruption of this balance can be associated with high 'exploration' behavior, because high loss may be related to impaired learning during MW.

This may imply that participants may have difficulty learning the values of the symbols because of the changes in the reward, which may cause them to have less exploitation behavior (choices based on learned values). And so, this can lead to an imbalance between exploitation and exploration behaviors. This speculation brings forward two arguments: if there was a shift between the exploitation and exploration balance during MW towards more exploration behavior (thus less exploitation), it might have happened due to participants engaged in more exploration during MW. Or possibly, it could have happened because the exploration tendency of participants was not influenced by MW, hence leading to impaired learning, which caused reduced exploitation behavior. This is because if a participant does not process the information effectively, it could lead to impaired task performance, hence potentially resulting in reduced exploitation behavior. To argue this, we can refer to our pupillometry data because the reduced reward processing can indicate impaired outcome processing and hence, lead to less optimal learning, as we will have discussed further.

Regarding our third hypothesis, we expected longer RTs during MW. Conversely, the results we obtained showed the opposite effect and did not confirm our hypothesis because the participants showed faster and more impulsive behavior instead of longer RTs. This situation can be explained in two ways; this hypothesis was based on the argument that participants while experiencing MW, would have their attention diverted from the task (i.e., perceptual decoupling), which might result in a longer response time simply because their attention not being on the task. However, the results obtained did not coincide with our hypothesis, instead, we found that participants responded faster, and impulsively when they experienced MW. We assume the reason for this situation was that, due to the nature of RL tasks, participants had to respond within a certain time window, otherwise they would face a time alert (i.e., 'Timeout' emoticon). This suggests that the time alert did not allow participants to be overly sluggish with responding while in the MW condition (i.e., the task environment did not enable extended RTs), on the contrary, participants were required to respond to the task within a certain time window. This might be because participants might have a less optimal estimation of the time available for making choices. Therefore, this may possibly have made participants feel more pressure to make decisions quickly, resulting in more impulsive and faster RTs during MW.

Furthermore, we examined how the variability of RTs (represented by the RTCV variable) was related to the frequency of MW reports because it has been shown to be an objective behavioral marker of MW (Bastian & Sackur, 2013). However, our results did not support the idea that RTCV is a predictor of MW in the RL context, which was in contrast to what has been found in sustained attention tasks (e.g., SART, Go/No-Go) in the literature (Maillet et al., 2020). We can interpret the reason for this situation as follows. Firstly, our RL task may not have been sensitive enough to examine the association between RT variability and MW. This is because, participants had to respond within a fixed 2-second window, perhaps if this time window had been extended, for instance, to 3 seconds, their RTs might have varied more. This would have allowed the participants to be more sluggish with their responses. Therefore, adding more time for responding could potentially indicate that RTCV would have been a predictor of MW. The second reason was the possibility that the association between RTCV as a behavioral measure of MW may not manifest across different contexts and cognitive tasks.

In other words, the reward- and loss-based nature of RL, as well as the limited response time, may have potentially affected reaction times, making response time differences difficult to capture. Involving motivational components in such tasks (i.e., RL), might change RT by influencing neural mechanisms associated with reacting to emotionally salient stimuli, such as dopamine (DA)(Aston-Jones & Cohen, 2005). The role of DA in the reinforcement learning system has been shown in a fair amount of studies by now (Lindsey & Litwin-Kumar, 2022; Rios et al., 2023; Varazzani et al., 2015). Notably, Montague (2004) posits that DA plays a role in better estimation of when rewards will emerge, hence providing more optimal behavior while reward-seeking (Montague et al., 2004). This is because the phasic release of DA can influence

RT variability by leading to action invigoration (Aston-Jones & Cohen, 2005). Hence, we can assume that as a result of releasing DA, and thus the reward-related increase in RT could potentially overshadow the effects of MW in the RL. So, RTCV being a predictor for MW may be specific to certain cognitive tasks, such as SART (Liu et al., 2023) and finger tapping (Groot et al., 2022) rather than reward and loss-based reinforcement learning tasks.

First of all, the underlying reason for our hypothesis about the RTs was the idea that relatively longer RTs during the task may also indicate a slow and unfocused state that leads to poor performance, but our results showed the opposite. Although the results were not consistent with our hypothesis about RTs, we speculated about why we found shorter RTs rather than longer during MW. We suggest that might be because both short and long RTs may result from opposite factors. This means that; for example, when short RTs are combined with high task performance, it may lead to more efficient information processing and optimal choice behavior. However, we did not observe this effect because choice suboptimality was higher in the MW condition in our study. Generally, short RTs may also index impulsivity and suboptimality, as in many cases of cognitive tasks, which is consistent with our results because short RTs led to impulsive and suboptimal behaviors which were associated with low task performance in the RL task.

In our final hypothesis concerning the behavioral measurements, we examined how motivation might have been shaped by the reward rate in the RL task and how it was related to participants' propensity of MW during the RL task. Although we did not have a parameter that directly measured motivation in the RL task, we assumed that average reward rates prior to each thought probe were related to the value participants attached to the task and their intrinsic motivation. In line with our results, we found the average reward rate was negatively associated with MW as we anticipated. At one point, this is compatible with the current concern hypothesis (Klinger, 1971; Randall et al., 2014), because the participants' higher average reward rates can indicate that they were motivated towards the RL task and considered it more valuable/interesting than their current concerns, thus they were more focused and engaged on the task.

Therefore, we can interpret this as follows: the current concerns of participants with higher average rewards were no more valuable or rewarding than the RL task, and in this case, their minds were less likely to generate potentially task-unrelated thoughts. To enhance this argument, we can exemplify it through the study by Padmala and Pessoa (2011) where the authors investigated how motivation influences cognitive control, especially in challenging situations. They proposed that when participants were motivated to receive rewards, they controlled challenging situations better compared to no-reward conditions. They also suggest that when participants expected a reward, they observed stronger reward-related activity in the attention-related frontoparietal regions of the brain (Padmala & Pessoa, 2011). In line with these findings, it is possible to say that reward-related motivation (Schultz, 2015) can enhance cognitive control in challenging task situations, and this can also be associated with better task performance.

We can explain this through the executive failure hypothesis (Kane & McVay, 2012; McVay & Kane, 2010), as executive failure suggests that MW occurs during cognitive tasks because the executive control system may fail to filter out task-irrelevant thoughts. If studies suggest that reward-related activity increases activity in the prefrontal cortex, this may potentially increase activity-related motivation in the executive control system and therefore result in less MW reporting in the RL task. The study conducted by (Kawagoe, 2022) suggested that the relationship between motivation and MW at the trait level is controlled by the executive system, and as a result, Kawagoe (2022) observed that spontaneous MW occurs especially in situations with lower levels of motivation, whose relationship relies on executive function. In this regard, it is possible to explain the relationship between average reward and

MW in this way, but still, since our RL task was not a task that directly measures motivation, we can consider this as a limitation. Because we do not know whether the average reward in the RL was directly related to participants' intrinsic motivation.

Finally, we included an exploratory model to compare the effects of performance measures Loss and the Average reward together. The reason behind this is that the Loss variable was not a direct measure of choice optimality because it depends on random fluctuations of the reward as well as on making the right decisions. By including both measures, we controlled the difference in the reward and focused on the optimality of choices. We expected that both the Loss and average reward variables would remain significant in our hypothesis, however, the average reward variable did not remain significant. This indicates that MW was mainly influencing performance through the Loss variable, and reward magnitude was not important enough when we control for overall performance. If we elaborate on this further: when participants perform worse, it would resulting high loss, indicating suboptimality of choices. But another factor that we need to take into consideration is that the Loss variable was also confounded by other factors such as task difficulty, which actually depends on the current value interval of the random walk. So easy trials have large value differences while difficult trials have small differences between the stimuli. Therefore, we can assume that if there is a high loss for any participant, it can be due to relatively suboptimal choices on easy trials, or many suboptimal choices on difficult trials. This is because we did not disentangle choice optimality from choice difficulty. Nevertheless, the Loss variable caught how many points participants lost because they did not choose the better option, and that was positively related to MW. Therefore, this indicates that if we consider performance through the Loss variable, the overall reward magnitude was not really effective. As a result, even if we have some indication of a relationship between reward rate and MW, it was not convincing enough because its effect disappeared if we also controlled for overall performance through the Loss variable. We clarified that choice suboptimality was robustly related to MW and hence, MW has a negative impact on RL performance. However, when we looked at the relationship between reward rate, motivation, and MW, we could not reach a convincing enough result. In this regard, we can suggest that what was important for RL performance was not the magnitude of the reward, but how people performed during the RL task.

Taking all the behavioral analyses together, we can argue that MW was less associated with average reward (which was related to motivation) than task performance (choice optimality reflected by the Loss variable and impulsivity reflected by mean RTs).

Pupillometry analysis

The results of the pupillometry analysis showed us that the pupillary responses in the OT condition were larger than in the MW condition, and accordingly, sensory/perceptual processing might have been more efficient in the OT condition compared to the MW condition. Therefore, it is possible to interpret this in two ways: First, the attenuated pupillary response in the MW condition can reconciled with the perceptual decoupling hypothesis of MW, because it indicates that stimuli were processed less in the MW than in the OT condition. However, a second argument is that the pupillary response to the onset of stimuli may be connected to the decision-making process. This means that when participants were presented with symbols and had to choose between them, the mental effort involved in making that decision could also potentially influence their pupils' response. This is because it has been shown through various cognitive tasks depending on executive control and decision-making, that pupil size is strongly associated with mental effort (Alnæs et al., 2014; Takeuchi et al., 2011). In either approach, we expected weaker pupil response in the MW condition and larger pupil response in the OT condition, although it is somewhat difficult to clarify whether this difference was due to perceptual decoupling, mental effort, or both. Perhaps new studies will provide a new perspective on interpreting the discrepancies in pupil responses between sensory and reward processing.

While this discussion was from the perspective of pupillary responses for processing the stimuli, we also need to consider how pupils respond to the reward process. The results we obtained for reward processing were consistent with stimulus processing. That is, we again observed that the pupillary responses in the MW condition were weaker in reward processing than in the OT condition. The reason for this is generally similar to the perspectives we discussed in stimulus processing, however, in reward processing, the relationship of stimulus values with motivation, and the learning process should also be discussed. That is because it has been widely known that the pupil responds to emotional stimuli more vigorously than to emotionally neutral stimuli, which means pupil responses are sensitive to reward and loss (Aston-Jones & Cohen, 2005; Joshi et al., 2016; Pulcu & Browning, 2017).

We can interpret this hypothesis as follows: we primarily expected that the reward-related pupil responses in the OT condition would be larger than in the MW condition. The reason for this is that pupillary responses in stimulus processing were weak in the MW condition, which also indicates that reward-related responses were accordingly weak. It is possible to explain this situation again with the perceptual decoupling hypothesis because participants were not paying attention to the stimulus in the MW condition. Therefore, they were not paying attention to the reward-related feedback because their attention was not on the task anyway, resulting in weak pupillary responses.

Another interpretation that can be made for the distinction between MW and OT is, that the larger reward-related pupil responses in the OT condition may be related to motivation as we mentioned at the beginning. This suggests that rewards have an emotional salience that influences pupil responses. In other words, if the participant has a high reward, it indicates that the pupillary responses would be larger because, in the case of receiving a high reward, this could potentially make the participant excited about the reward, perhaps triggering the releasing of more dopamine (Aston-Jones & Cohen, 2005; Berridge & Waterhouse, 2003) and thus the pupil gets larger. It is important to note that the reward-processing phase of the pupils is crucial in the context of the RL mechanism because it demonstrates that the rewards are learned (Sutton & Barto, 2014). Therefore, successful reward processing indicates that the participant can perform better in the next trial because the participant has better-processed information. This approach is actually related to how optimal the outcomes that guide future learning are (Sutton & Barto, 2014), and this is practically the basis of reinforcement learning. On the contrary, since the stimulus was less processed, the pupillary reactions to the reward were also less in the MW condition, and thus, the insufficient processing of the reward logically influenced the learning process negatively and as a result, led to poor performance.

Taken together, both of our hypotheses regarding pupil measurements were confirmed according to the results, and we can argue that the pupil in the MW condition responded weaker in both the stimulus and reward processing phase compared to the OT condition, which was potentially negatively affecting task performance.

Strengths, Limitations and Future Directions

Participants

This laboratory-based study has enhanced statistical power thanks to having a sample size of 50 participants. In addition, many participants did not have experience in such an experiment setting before and were not familiar with such cognitive tasks, which helped to minimize the effects of potential biases towards to task.

Additionally, forthcoming investigations may consider establishing more comprehensive participant inclusion criteria considering the sensitivity of the eye-tracker: such as laser surgery, dry eye (i.e., xerophthalmia), or watery eye (i.e., epiphora).

Task Structure

We have comprehensively discussed many potential limitations within the RL task structure in the discussion. To summarize these, first of all, the RL task could not directly measure motivation can be a limitation of this study. Although we consider motivation with the average reward obtained in the task, we cannot predict the reason for this with certainty. This is due to our exploratory model, where we controlled both Loss and Average reward variables, we found that Average reward did not have a significant effect on MW. In this respect, future studies involving RL could integrate a different parameter (e.g., self-reported motivation scales) to measure the potential effects of motivation on task performance.

Another possible limitation regarding our RL task is that having the response time to the symbols in the task within a time frame may have prevented the variability of the RTs, thus increasing the sensitivity of looking at the putative RTCV and MW. Because the limited response time has reduced the variability of the RTs. In this respect, it may be due to the fact that the RL task concept does not have the necessary sensitivity to measure RTCV or a response pattern like other cognitive tasks (e.g., SART, Go/No-Go).

Pupillometry

As we observed from the pupillometry analysis, there was a difference between the pupil responses in both stimuli and reward processing for MW and OT conditions. However, there is still uncertainty about whether this difference in the MW and OT conditions stems from perceptual decoupling or mental effort. Even though it is reasonable to say that both arguments are logical from different aspects, perhaps future research will provide us with a different angle. For instance, the reward might be included in the trial with a delay, thus pupil response to the stimuli is concluded once the reward processing has begun.

Conclusion

In this study, the potential effects of MW on RL were investigated, and MW was evidently found to have negative effects on the RL task as in many other cognitive tasks. The behavioral findings were also synthesized with pupillometry analysis, and the results are comprehensively discussed. Our research findings contribute to the deeper understanding of the MW dynamics by including the RL paradigm in this study, in addition to the traditionally used cognitive tasks in MW studies. Furthermore, methodological problems and their potential effects were mentioned in this study, and the importance of investigating the questions left open for future studies was emphasized.

References

Aston-Jones, G., & Cohen, J. D. (2005). AN INTEGRATIVE THEORY OF LOCUS COERULEUS-NOREPINEPHRINE FUNCTION: Adaptive Gain and Optimal Performance. *Annual Review of Neuroscience*, *28*(1), 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Baird, B., Smallwood, J., & Schooler, J. W. (2011). Back to the future: Autobiographical planning and the functionality of mind-wandering. *Consciousness and Cognition*, *20*(4), 1604–1611. https://doi.org/10.1016/j.concog.2011.08.007

Bonifacci, P., Viroli, C., Vassura, C., Colombini, E., & Desideri, L. (2023). The relationship between mind wandering and reading comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *30*(1), 40–59. https://doi.org/10.3758/s13423-022-02141-w

Brocher, A., & Graf, T. (2017). Decision-related factors in pupil old/new effects: Attention, response execution, and false memory. *Neuropsychologia*, *102*, 124–134. https://doi.org/10.1016/j.neuropsychologia.2017.06.011

Carriere, J. S. A., Cheyne, J. A., & Smilek, D. (2008). Everyday attention lapses and memory failures: The affective consequences of mindlessness. *Consciousness and Cognition*, *17*(3), 835–847. https://doi.org/10.1016/j.concog.2007.04.008

Cohen, R. A. (2014). *The Neuropsychology of Attention*. Springer US. https://doi.org/10.1007/978-0-387-72639-7

Esterman, M., & Rothlein, D. (2019). Models of sustained attention. *Current Opinion in Psychology*, *29*, 174–180. https://doi.org/10.1016/j.copsyc.2019.03.005

Groot, J. M., Csifcsák, G., Wientjes, S., Forstmann, B. U., & Mittner, M. (2022). Catching wandering minds with tapping fingers: Neural and behavioral insights into task-unrelated cognition. *Cerebral Cortex*, *32*(20), 4447–4463. https://doi.org/10.1093/cercor/bhab494

Jackson, J. D., & Balota, D. A. (2012). Mind-wandering in Younger and Older Adults: Converging Evidence from the Sustained Attention to Response Task and Reading for Comprehension. *Psychology and Aging*, *27*(1), 106–119. https://doi.org/10.1037/a0023933

Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. *Neuron*, *89*(1), 221–234. https://doi.org/10.1016/j.neuron.2015.11.028

Kam, J. W. Y., Dao, E., Farley, J., Fitzpatrick, K., Smallwood, J., Schooler, J. W., & Handy, T. C. (2011). Slow Fluctuations in Attentional Control of Sensory Cortex. *Journal of Cognitive Neuroscience*, *23*(2), 460–470. https://doi.org/10.1162/jocn.2010.21443

Kane, M. J., & McVay, J. C. (2012). What Mind Wandering Reveals About Executive-Control Abilities and Failures. *Current Directions in Psychological Science*, *21*(5), 348–354. https://doi.org/10.1177/0963721412454875

Kawagoe, T. (2022). Executive failure hypothesis explains the trait-level association between motivation and mind wandering. *Scientific Reports*, *12*(1), 5839. https://doi.org/10.1038/s41598-022-09824-3

Killingsworth, M. A., & Gilbert, D. T. (2010). A Wandering Mind Is an Unhappy Mind. *Science*, *330*(6006), 932–932. https://doi.org/10.1126/science.1192439

Klinger, E., Gregoire, K. C., & Barta, S. G. (1973). Physiological Correlates of Mental Activity: Eye Movements, Alpha, and Heart Rate During Imagining, Suppression, Concentration, Search, and Choice. *Psychophysiology*, *10*(5), 471–477. https://doi.org/10.1111/j.1469-8986.1973.tb00534.x

Kvavilashvili, L., & Mandler, G. (2004). Out of one's mind: A study of involuntary semantic memories. *Cognitive Psychology*, *48*(1), 47–94. https://doi.org/10.1016/S0010-0285(03)00115-4

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A Window to the Preconscious? *Perspectives on Psychological Science*, *7*(1), 18–27. https://doi.org/10.1177/1745691611427305

Leary, M. R., & Hoyle, R. H. (2013). *Handbook of Individual Differences in Social Behavior*. Guilford Publications. https://books.google.no/books?id=67xcAgAAQBAJ

Lindsey, J., & Litwin-Kumar, A. (2022). Action-modulated midbrain dopamine activity arises from distributed control policies. *Advances in Neural Information Processing Systems*, *35*, 5535–5548.

Liu, S., Rabovsky, M., & Schad, D. J. (2023). Spontaneous mind wandering impairs model-based decision making. *PLOS ONE*, *18*(1), e0279532. https://doi.org/10.1371/journal.pone.0279532

Mathôt, S. (2013). *A simple way to reconstruct pupil size during eye blinks*. https://doi.org/10.6084/m9.figshare.688001.v1

Mathot, S. (2018). *Pupillometry: Psychology, Physiology, and Function* (1). *1*(1), Article 1. https://doi.org/10.5334/joc.18

Mathôt, S., & Vilotijević, A. (2022). Methods in cognitive pupillometry: Design, preprocessing, and statistical analysis. *Behavior Research Methods*, *55*(6), 3055–3077. https://doi.org/10.3758/s13428-022-01957-7

McVay, J. C., & Kane, M. J. (2010). Does mind wandering reflect executive function or executive failure? Comment on Smallwood and Schooler (2006) and Watkins (2008). *Psychological Bulletin*, *136*(2), 188–197. https://doi.org/10.1037/a0018298

Mittner, M. (2020). pypillometry: A Python package for pupillometric analyses. *Journal of Open Source Software*, *5*(51), 2348. https://doi.org/10.21105/joss.02348

Montague, P. R., Hyman, S. E., & Cohen, J. D. (2004). Computational roles for dopamine in behavioural control. *Nature*, *431*(7010), 760–767. https://doi.org/10.1038/nature03015

Padmala, S., & Pessoa, L. (2011). Reward Reduces Conflict by Enhancing Attentional Control and Biasing Visual Cortical Processing. *Journal of Cognitive Neuroscience*, *23*(11), 3419–3432. https://doi.org/10.1162/jocn_a_00011

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pulcu, E., & Browning, M. (2017). Affective bias as a rational response to the statistics of rewards and punishments. *eLife*, *6*, e27879. https://doi.org/10.7554/eLife.27879

Randall, J. G., Oswald, F. L., & Beier, M. E. (2014). Mind-wandering, cognition, and performance: A theory-driven meta-analysis of attention regulation. *Psychological Bulletin*, *140*(6), 1411–1431. https://doi.org/10.1037/a0037428

Reichle, E. D., Reineberg, A. E., & Schooler, J. W. (2010). Eye Movements During Mindless Reading. *Psychological Science*, *21*(9), 1300–1310. https://doi.org/10.1177/0956797610378686

Rios, A., Nonomura, S., Kato, S., Yoshida, J., Matsushita, N., Nambu, A., Takada, M., Hira, R., Kobayashi, K., Sakai, Y., Kimura, M., & Isomura, Y. (2023). Reward expectation enhances action-related activity of nigral dopaminergic and two striatal output pathways. *Communications Biology*, *6*(1), 1–14. https://doi.org/10.1038/s42003-023-05288-x

Robison, M. K., & Unsworth, N. (2018). Cognitive and contextual correlates of spontaneous and deliberate mind-wandering. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(1), 85–98. https://doi.org/10.1037/xlm0000444

Schooler, J., Reichle, E., & Halpern, D. V. (2004). Zoning-out during reading: Evidence for dissociations between experience and meta-consciousness. *Thinking and Seeing: Visual Metacognition in Adults and Children*, 204–226.

Schooler, J. W. (2002). Re-representing consciousness: Dissociations between experience and meta-consciousness. *Trends in Cognitive Sciences*, *6*(8), 339–344. https://doi.org/10.1016/S1364-6613(02)01949-6

Schubert, A.-L., Frischkorn, G. T., Sadus, K., Welhaf, M., Kane, M. J., & Rummel, J. (2023). *The brief mind wandering three-factor scale (BMW-3)* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/cy9rx

Schultz, W. (2015). Neuronal Reward and Decision Signals: From Theories to Data. *Physiological Reviews*, *95*(3), 853–951. https://doi.org/10.1152/physrev.00023.2014

Seli, P., Carriere, J. S. A., & Smilek, D. (2015). Not all mind wandering is created equal: Dissociating deliberate from spontaneous mind wandering. *Psychological Research*, *79*(5), 750–758. https://doi.org/10.1007/s00426-014-0617-x

Semendeferi, K., Lu, A., Schenker, N., & Damasio, H. (2002). Humans and great apes share a large frontal cortex. *Nature Neuroscience,* *5*(3), 272–276. https://doi.org/10.1038/nn814

Smallwood, J., Nind, L., & O’Connor, R. C. (2009). When is your head at? An exploration of the factors associated with the temporal focus of the wandering mind. *Consciousness and Cognition*, *18*(1), 118–125. https://doi.org/10.1016/j.concog.2008.11.004

Smallwood, J., & O’Connor, R. C. (2011). Imprisoned by the past: Unhappy moods lead to a retrospective bias to mind wandering. *Cognition and Emotion*, *25*(8), 1481–1490. https://doi.org/10.1080/02699931.2010.545263

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, *132*(6), 946–958. https://doi.org/10.1037/0033-2909.132.6.946

Stawarczyk, D., Majerus, S., Maj, M., Van Der Linden, M., & D’Argembeau, A. (2011). Mind-wandering: Phenomenology and function as assessed with a novel experience sampling method. *Acta Psychologica*, *136*(3), 370–381. https://doi.org/10.1016/j.actpsy.2011.01.002

Sutton, R. S., & Barto, A. G. (2014). *Reinforcement Learning: An Introduction*.

Thomson, D. R., Seli, P., Besner, D., & Smilek, D. (2014). On the link between mind wandering and task performance over time. *Consciousness and Cognition*, *27*, 14–26. https://doi.org/10.1016/j.concog.2014.04.001

Vallender, E. J., Mekel-Bobrov, N., & Lahn, B. T. (2008). Genetic basis of human brain evolution. *Trends in Neurosciences*, *31*(12), 637–644. https://doi.org/10.1016/j.tins.2008.08.010

Varazzani, C., San-Galli, A., Gilardeau, S., & Bouret, S. (2015). Noradrenaline and Dopamine Neurons in the Reward/Effort Trade-Off: A Direct Electrophysiological Comparison in Behaving Monkeys. *The Journal of Neuroscience*, *35*(20), 7866–7877. https://doi.org/10.1523/JNEUROSCI.0454-15.2015

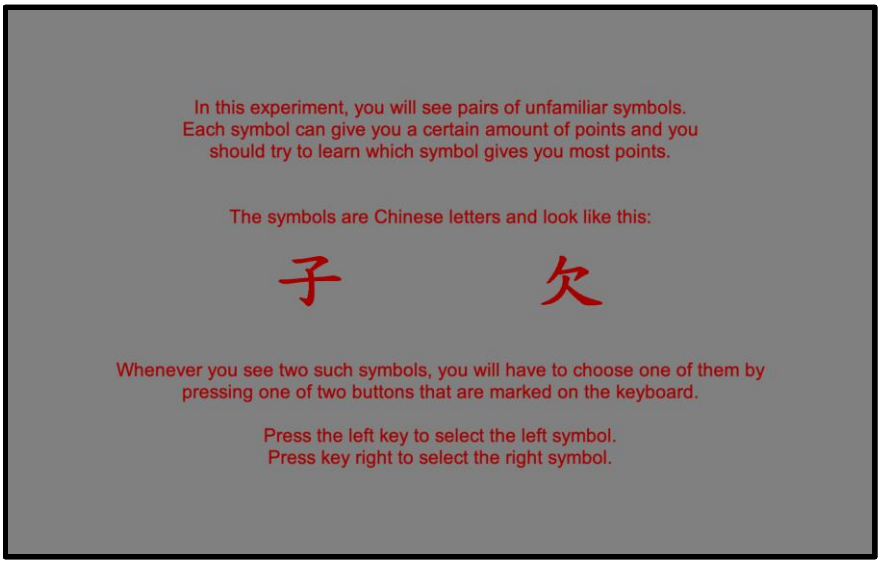
Vazey, E. M., Moorman, D. E., & Aston-Jones, G. (2018). Phasic locus coeruleus activity regulates cortical encoding of salience information. *Proceedings of the National Academy of Sciences*, *115*(40). https://doi.org/10.1073/pnas.1803716115

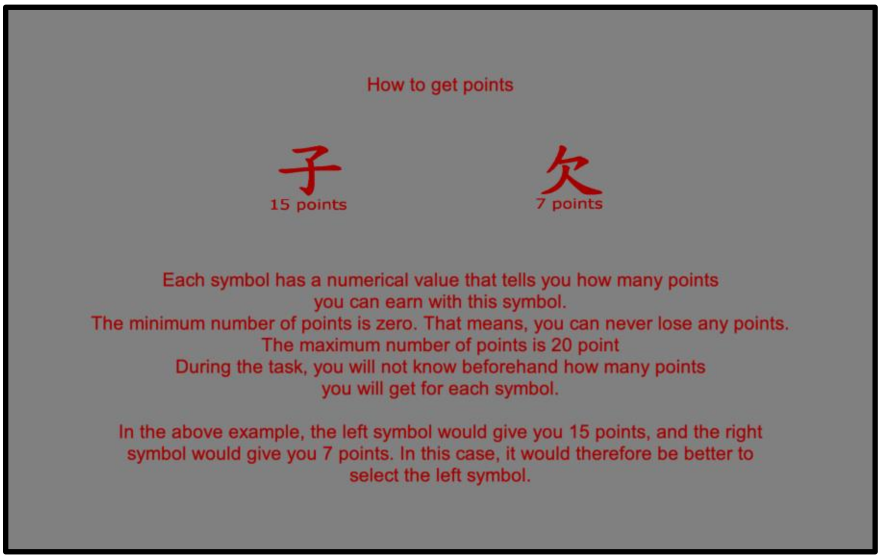
Verney, S. P., Granholm, E., & Dionisio, D. P. (2001). Pupillary responses and processing resources on the visual backward masking task. *Psychophysiology*, *38*(1), 76–83. https://doi.org/10.1111/1469-8986.3810076

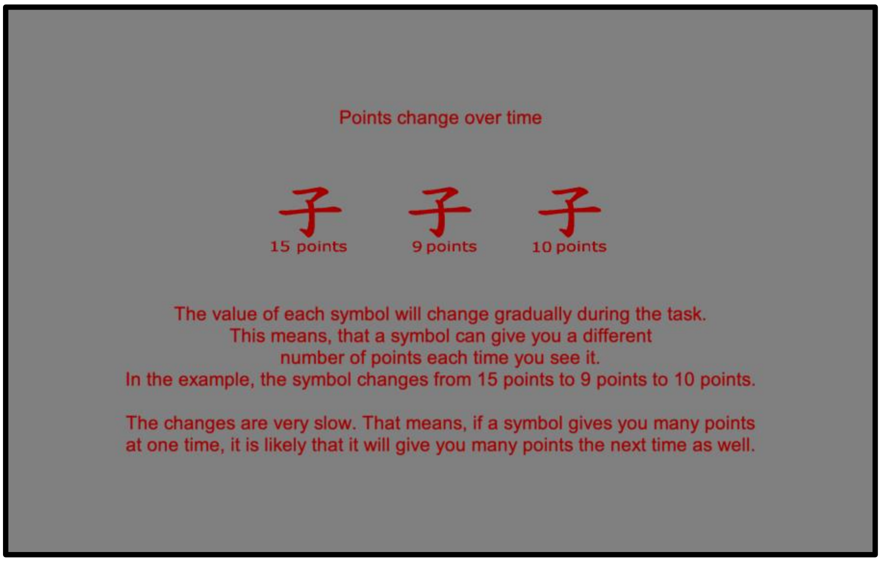
Wong, Y.-S., Willoughby, A. R., & Machado, L. (2023). Reconceptualizing mind wandering from a switching perspective. *Psychological Research*, *87*(2), 357–372. https://doi.org/10.1007/s00426-022-01676-w

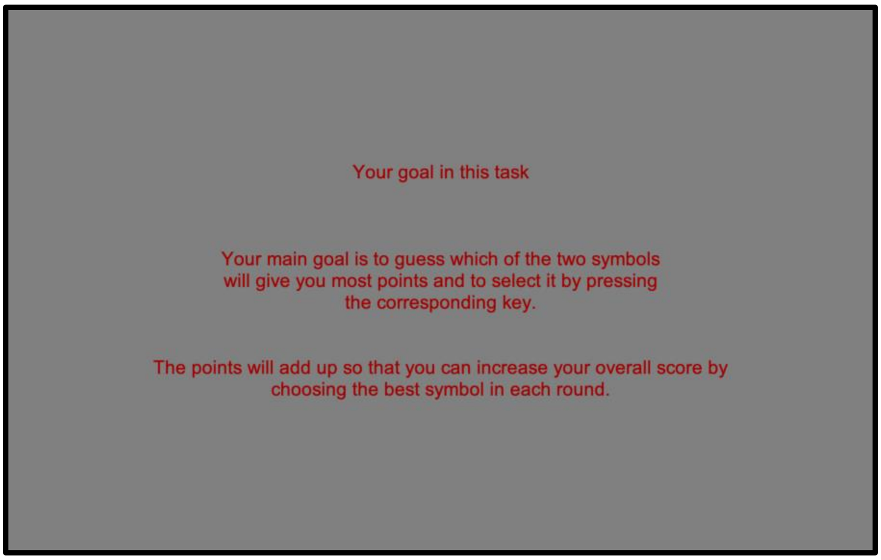
Appendix

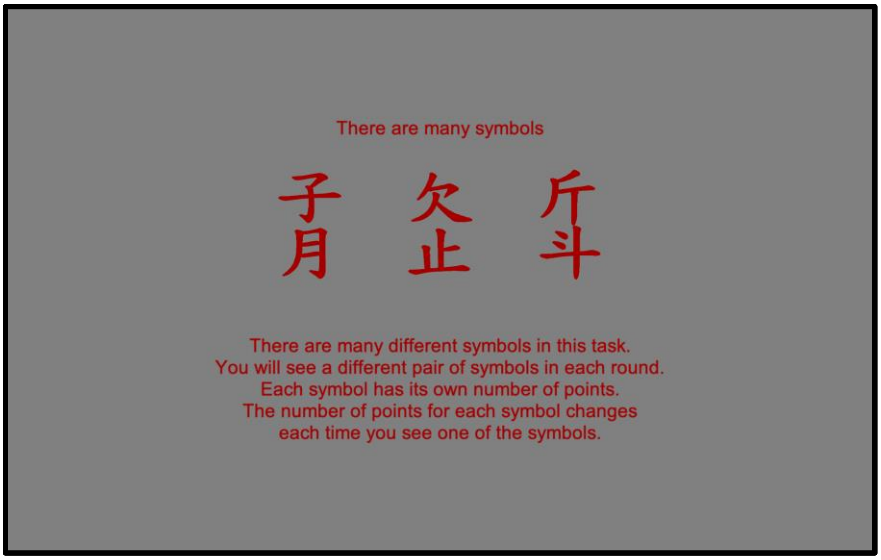
Task Instructions

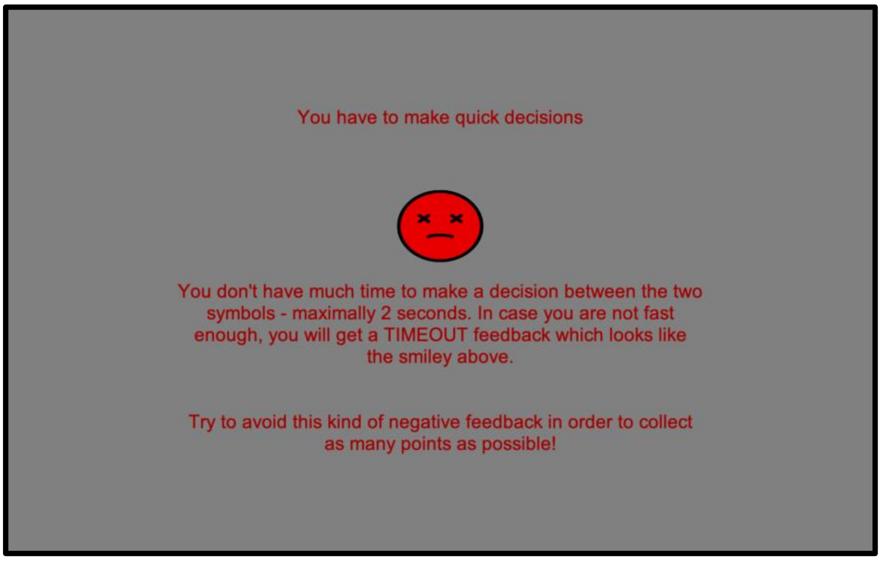


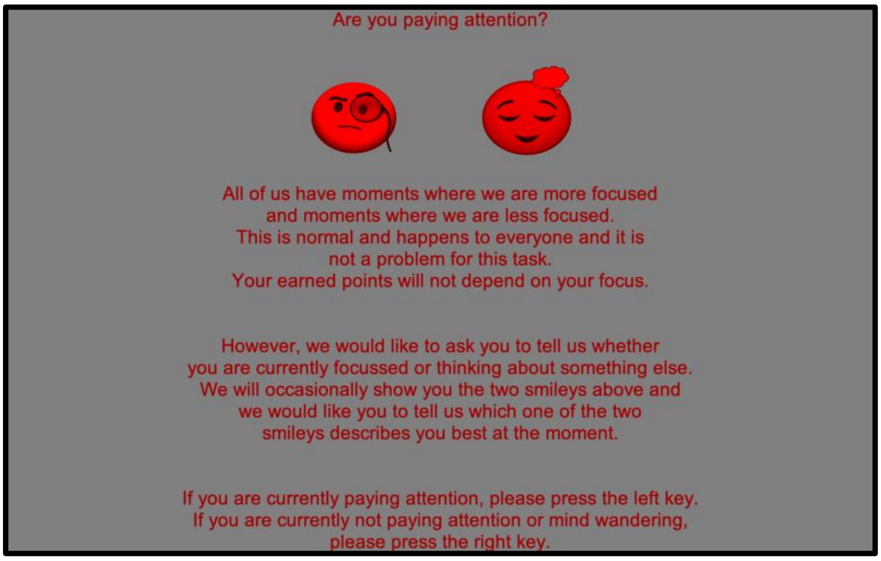


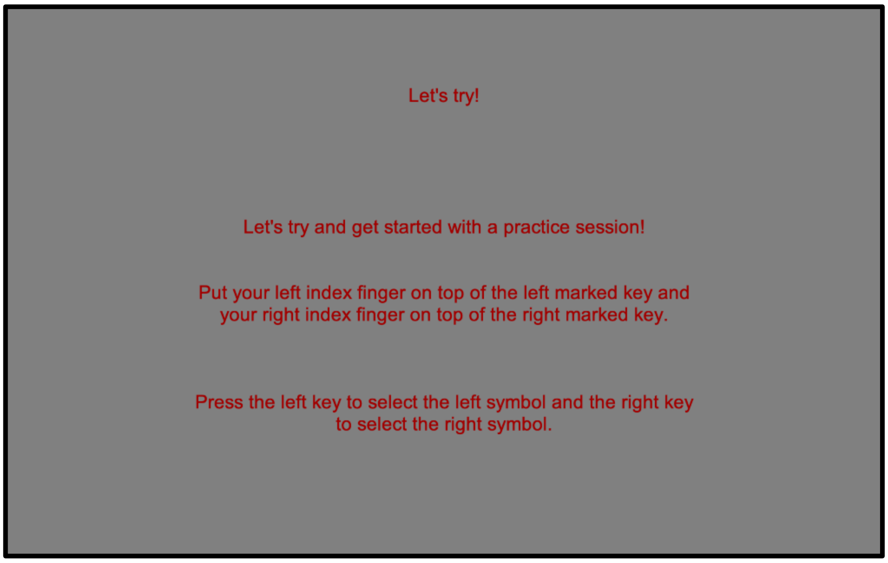

Points change over time
子 子 子
15 points
9 points
10 points
The value of each symbol will change gradually during the task.
This means, that a symbol can give you a different
number of points each time you see it.
In the example, the symbol changes from 15 points to 9 points to 10 points.
The changes are very slow. That means, if a symbol gives you many points
at one time, it is likely that it will give you many points the next time as well.


Your goal in this task
Your main goal is to guess which of the two symbols
will give you most points and to select it by pressing
the corresponding key.
The points will add up so that you can increase your overall score by
choosing the best symbol in each round.

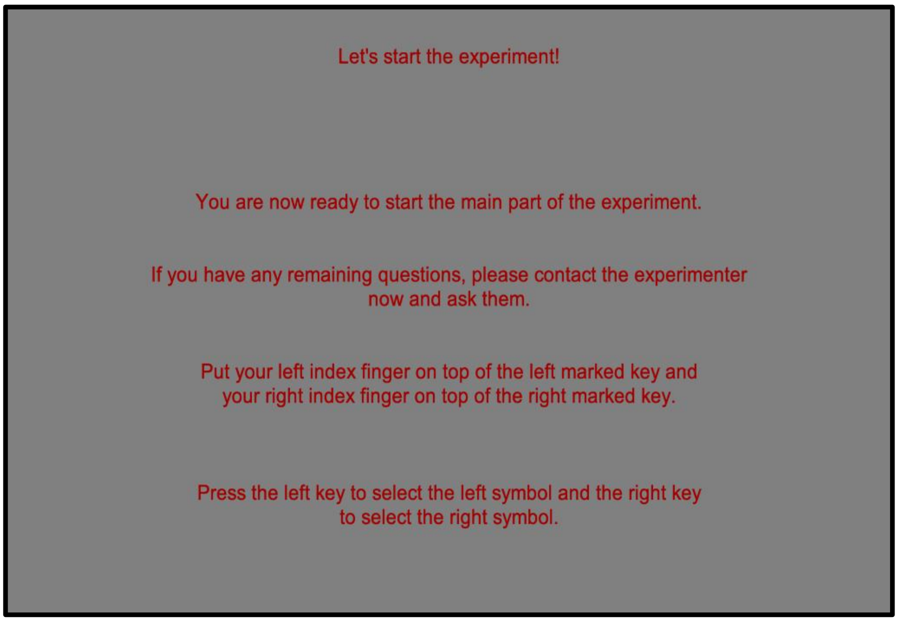

There are many symbols
子月
欠止
斤斗
There are many different symbols in this task.
You will see a different pair of symbols in each round.
Each symbol has its own number of points.
The number of points for each symbol changes
each time you see one of the symbols.


You have to make quick decisions
You don't have much time to make a decision between the two symbols - maximally 2 seconds. In case you are not fast enough, you will get a TIMEOUT feedback which looks like the smiley above.
Try to avoid this kind of negative feedback in order to collect as many points as possible!


Are you paying attention?
All of us have moments where we are more focused and moments where we are less focused.
This is normal and happens to everyone and it is not a problem for this task.
Your earned points will not depend on your focus.
However, we would like to ask you to tell us whether you are currently focussed or thinking about something else.
We will occasionally show you the two smileys above and we would like you to tell us which one of the two smileys describes you best at the moment.
If you are currently paying attention, please press the left key.
If you are currently not paying attention or mind wandering, please press the right key.


Let's try!
Let's try and get started with a practice session!
Put your left index finger on top of the left marked key and your right index finger on top of the right marked key.
Press the left key to select the left symbol and the right key to select the right symbol.


Thank you!
The training session is now over and we will start the main experiment soon.
Please make yourself comfortable so that you can complete the experiment without interruptions.
We will now calibrate the eye-tracker again and then start the experiment.
Press any key to continue.


Let's start the experiment!
You are now ready to start the main part of the experiment.
If you have any remaining questions, please contact the experimenter now and ask them.
Put your left index finger on top of the left marked key and your right index finger on top of the right marked key.
Press the left key to select the left symbol and the right key to select the right symbol.

Nettskjema

NFC

Username

NFC - Questionnaire

For each statement below, please indicate to which extent the statement is characteristic of you.
If the statement is extremely uncharacteristic of you (not at all like you), please tick "1" next to the question. If the statement is extremely characteristic of you (very much like you), please tick "5" next to the question. Of course, a statement can be neither very characteristic nor very uncharacteristic of you; if so, please use the number in the middle of the scale that describes the best fit.

1 - Very strongly disagree 5 - Very strongly agree

I would prefer complex to simple problems.

1
2
3
4
5

I like to have the responsibility of handling a situation that requires a lot of thinking.

1
2
3
4
5

Thinking is not my idea of fun.

1
2
3
4
5

I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.

1
2
3
4
5

I try to anticipate and avoid situations where there is likely a chance I will have to think in depth about something.

Nettskjema

1
2
3
4
5

I find satisfaction in deliberating hard and for long hours.

1
2
3
4
5

I only think as hard as I have to.

1
2
3
4
5

I prefer to think about small, daily projects to long-term ones.

1
2
3
4
5

I like tasks that require little thought once I've learned them.

1
2
3
4
5

The idea of relying on thought to make my way to the top appeals to me.

1
2
3
4
5

I really enjoy a task that involves coming up with new solutions to problems.

1
2
3

Nettskjema

4
5

Learning new ways to think doesn't excite me very much.

1
2
3
4
5

I prefer my life to be filled with puzzles that I must solve.

1
2
3
4
5

The notion of thinking abstractly is appealing to me.

1
2
3
4
5

I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.

1
2
3
4
5

I feel relief rather than satisfaction after completing a task that required a lot of mental effort.

1
2
3
4
5

It's enough for me that something gets the job done; I don't care how or why it works.

1
2
3
4
5

Nettskjema

I usually end up deliberating about issues even when they do not affect me personally.

1
2
3
4
5

Nettskjema

BMW-3 - Questionnaire 2

Username

Brief Mind Wandering Three-Factor Scale (BMW-3)

The several statements below describe common mental states in everyday life. Please read each statement carefully and decide how much you agree or disagree with that statement. Please note that there are no right or wrong responses, and please respond based on your intuition.

0 = fully disagree
1 = somewhat disagree
2 = neutral
3 = somewhat agree
4 = fully agree

While listening to a presentation, my thought start to trail off unintentionally.

0
1
2
3
4

When watching TV, other things inadvertently cross my mind.

0
1
2
3
4

I am often absentminded.

0
1
2
3
4

When I am engaged in an activity, my thoughts wander to other things all by themselves.

0
1
2
3
4

I make my thoughts wander so that time passes faster.

Nettskjema

0
1
2
3
4

I deliberately allow my mind to wander to escape the daily grind.

0
1
2
3
4

I distract myself in monotonous situations by letting my thoughts run free.

0
1
2
3
4

I actively use the time during routine tasks to mull over other things in the meanwhile.

0
1
2
3
4

It takes a very long time for me to notice that my thoughts have wandered off.

0
1
2
3
4

I quickly catch myself when I am not listening attentively.

0
1
2
3
4

It takes me a while before I realize that I zoned out.

0
1
2

Nettskjema

3
4

I immediately notice when my thoughts are not in the here and now.

0
1
2
3
4