It’s in the Eyes: Memory Effects of Preparation Reflected in the Pupil

Sarah van den Bosch

S3928756

s.m.van.den.bosch@student.rug.nl

BCN Research Master, C-track

Faculty of Science and Engineering

University of Groningen

Supervisor:

Wouter Kruijne

w.kruijne@rug.nl

Department of Experimental Psychology

Faculty of Behavioral and Social Sciences

University of Groningen

Abstract

A warning preceding an event improves response speed, demonstrating the ability to prepare a response in advance. The multiple trace theory of temporal preparation suggests that memory traces influence preparation based on past experiences. Previous research has found that this memory-guided preparation can be momentarily disrupted, indicating that trace retrieval does not affect each trial equivocally. One hypothesized candidate mechanism is that in cases of high arousal, participants' preparation is driven by explicit task goals, rather than memories retrieved from past trials. In the present study, we utilized pupil dilation as a measure for arousal, and assessed whether pupil dilation is a predictor of memory effects in preparation. We found that baseline pupil size was large at moments of high arousal, particularly during the disruption of memory-guided preparation. Furthermore, we found that the mean transient pupil response reflected differential preparation, resulting from participants' learned associations. Our results demonstrate a link between the pupil and behavior, providing valuable insights into the role of arousal in memory effects during temporal preparation.

Key words: Timing, pupillometry, temporal preparation, learning, arousal

Introduction

Temporal preparation and memory

Timing plays a vital role in our everyday activities. We apply timing to virtually everything, whether we are navigating busy traffic, timing the double-click of a computer mouse, or timing our input in conversation as to not interrupt others. Across all these activities, an intriguing cognitive skill emerges—one that involves preparing for the temporal demands of upcoming events. Notably, research has shown that a warning allows us to respond faster; i.e., we respond faster to "ready… go!" than to "go!" on its own (Niemi & Näätänen, 1981; Woodrow, 1914). This indicates that some form of preparation takes place between the warning and the response cue that facilitates a faster response. This preparation in anticipation of a future event is known as *temporal preparation*.

Temporal preparation is commonly studied using the variable foreperiod paradigm. At its core, this experimental paradigm is a reaction time task: participants are asked to respond to a target stimulus (S2) as fast as possible. S2 is preceded by a warning stimulus (S1), and the time between S1 and S2 is called the foreperiod (FP). The variable foreperiod paradigm derives its name from the between-trial variation in FP length. A common manipulation in this paradigm is to vary the distribution of these FPs across blocks or across participants (e.g. Los et al., 2017; Näätänen, 1971; Niemi & Näätänen, 1981; Trillenberg et al., 2000). Depending on the condition of the experimental block, a trial's FP can be sampled from either a uniform distribution or a biased distribution. When FPs are sampled from a uniform distribution, they have equal probability of being long or short, whereas biased distributions show a bias towards either longer or shorter FPs.

Varying the FP in this way produces various effects (Los et al., 2014; Salet et al., 2022). The simplest effect is that mean reaction time (RT) decreases as the FP duration increases (Los et al., 2001). Furthermore, preparation is dependent on the FP distribution, i.e., a participant responds faster to short intervals if short intervals are more prevalent (Niemi & Näätänen, 1981; Zahn & Rosenthal, 1966). Third, sequential effects describe the effect of the previous trial(s); even if participants try to keep track of the complete distribution, they still heavily rely on what occurred on the preceding trials to inform their next preparation (Los & Agter, 2005; Los & van de Heuvel, 2001). Fourth, transfer effects are long-lasting effects of the previous experimental blocks, which are visible even up to a week later (Crowe et al., 2019; 2021; Los et al., 2017; Mattiesing et al., 2017). Specifically, biased distributions will continue to affect response times long after having been changed to a uniform distribution (Los et al., 2021). These effects even occur when participants are explicitly informed that the distribution has changed (Kruijne et al., 2022; Los et al., 2017; 2021). Finally, associative cueing effects result from the associations participants form between the type of S1 and the distribution of the FP (Kruijne et al., 2022; Los et al., 2021). Participants can learn to predict the duration of the FP based on the type of S1 that precedes it. This acquired expectation leads to differential preparation, persisting after S1 loses its predictive value. Again, this effect also persist if participants are explicitly informed of the change in distributions.

A number of theories aim to explain the (neuro)cognitive mechanisms underlying aforementioned effects (see Los et al., 2014 for a review). The most straightforward of these is the hazard function (Janssen & Shadlen, 2005; Nobre & Correa, 2007; Trillenberg et al., 2000). The hazard function operates as follows: in a variable FP paradigm with four possible FP durations, there are four critical moments after the warning stimulus at which the target stimulus may appear. The hazard function describes the development of the conditional probability of the target stimulus appearing at each of these moments given that it has not yet

occurred. So, if there are four FPs, the chance of S2 appearing at the first critical moment is ¼. At the next critical moment, this chance increases to ⅓, with only three possible moments remaining. If S2 does not appear at the third possible moment, it has to appear at the last, increasing the probability to 1. The hazard model nicely explains the qualitative shape of the FP-RT function: the higher the hazard, the better the preparation thus the lower the RT. However, the link between hazard and the RT-FP function remains implicit (Los, 2013). Therefore, the hazard function is merely a statistical fact in its current state, rather than a valid explanatory construct. Furthermore, the hazard function fails to explain the other effects mentioned above, such as sequential or transfer effects. The hazard function assumes that participants simply tune into the new hazard function when the FP distribution changes, in order to optimally prepare. This leaves no room for the influence of previous trials on the current trial.

The multiple trace theory of temporal preparation (MTP) sought to address these issues by emphasizing the role of long-term memory in temporal preparation (Los et al., 2014). Multiple trace theories have three core assumptions: (1) All attended elements of an episode are stored in memory. (2) Each episode is stored as a single memory trace. (3) A new episode triggers the retrieval of memory traces that contain similar elements to those currently attended. In the context of this type of experimental paradigm, an episode is one trial. Every episode contains information about the balance between motor activation (A) and inhibition (I) during the trial. This A/I balance changes throughout the trial to facilitate fast responses: first, there is pre-activation of task-relevant effectors, which are inhibited during the FP. As soon as S2 appears, inhibition is released in order for a response to occur. The A and I values of an episode are stored as a memory trace. The central idea of MTP is that once S1 appears, it serves as a retrieval cue, and similar traces are retrieved, which affect the current preparatory state.

In multiple trace theories, recent traces have higher activation than earlier traces, providing a mechanism underlying sequential effects (Los et al., 2014). Furthermore, the model's trace retrieval mechanism dictates that old memory traces will continue to inform preparation in the transfer phase, as is visible in the data, demonstrating that transfer and cueing effects are supported by the MTP model (Kruijne et al., 2022; Los et al., 2021; Mattiesing et al., 2017).

A recent finding in the field of temporal preparation is the so-called "dip" in memory effects of preparation (Kruijne et al., 2022). The authors performed a rolling regression analysis on RT data in an acquisition-transfer experiment, in order to reveal how memory effects develop at a very fine timescale. The results show that the interaction coefficient (FP x S1 type) gradually grows during the acquisition phase, characteristic of a slowly acquired memory effect. However, the results reveal a dip that takes place immediately after participants are instructed that the FP-S1 contingency no longer holds. The authors surmised that the informed change from the acquisition phase to the transfer phase may have evoked a momentary suppression of automatic trace retrieval, either caused by or paired with increased arousal. This momentary suppression quickly decreases as participants revert back to learned associations.

This finding illustrates that memory effects can fluctuate. In particular, the dip seems to indicate that certain interventions can influence the way that learned associations guide preparation. The present paper served to further investigate the idea that fluctuation in memory effects is associated with arousal.

Pupil response, memory, and arousal

Our pupils dilate and constrict not only in response to environmental factors such as light, but cognitive factors as well (Laeng et al., 2012). It was first discovered that emotionally arousing stimuli evoke larger pupil sizes than non-relevant stimuli (Hess & Polt, 1960). Later, it was found that increasing memory load positively correlates with pupil size (Beatty & Kahneman, 1966). Recent developments in the field of visual perception have further advanced this idea of the relationship between the pupil and arousal/attention (e.g. Chatham et al., 2009; Laeng et al., 2011; Piquado et al, 2010).

Two types of pupil dilation can be distinguished: tonic pupil dilation and phasic pupil dilation. Tonic pupil dilation refers to the long-term changes in baseline pupil size and phasic pupil dilation refers to short-term transient pupil responses (Beatty, 1982). Research has found a relationship between large tonic pupil dilation and on-task reports, whereas smaller tonic pupil dilation is associated with mind-wandering (Gilzenrat et al., 2010; Unsworth & Robison, 2016; 2018). This suggests that tonic pupil size is lower at times of low task engagement, as apparent in poor behavioral performance (Aston-Jones et al., 1994; 2007).

Phasic (short-scale) pupillary changes correspond to the processing of task-relevant events, such as target detection (Privitera et al., 2008), as well as the anticipation of visual and auditory stimuli (Bradshaw, 1968; Knapen et al., 2016; Tressoldi et al., 2011). Pupils start dilating earlier and show higher relative levels of dilation when a target is expected after a shorter delay rather than a longer delay (Akdoğan et al., 2016). Furthermore, if uncertainty is low, such as in the case of fixed time intervals, arousal levels are low, reflected in smaller phasic pupillary changes compared to high uncertainty trials (Shalev & Nobre, 2022).

These findings may be explained by a link that has been discovered between pupil size and activity of the locus coeruleus (e.g. Gilzenrat et al., 2010; Murphy et al., 2014). The locus coeruleus (LC) is the hub of the LC-NE system, which is the primary source of norepinephrine in the brain (Berridge & Waterhouse, 2003). This system has widespread projections throughout the neocortex, including frontoparietal areas. Functions influenced by the LC-NE system include arousal, alertness, and readiness for action (Cohen et al., 2004). LC activity has been found to play an important role in regulating one's attentional state; specifically, LC plays a key role in focusing attention as well as disengaging ongoing action or thought (Aston-Jones & Cohen, 2005). Relevant for the present study is that a strong association has been found between LC neuron activity and pupil diameter (Koss, 1986), suggesting that LC may exert control over the eye muscles, so that LC activity gets reflected in the dilation of the pupil (Laeng et al., 2012). It should be acknowledged that this is a correlational relationship; a number of other regions have been found to play a role in pupil dilation as well (Nieuwenhuis et al., 2011; Wang & Munoz, 2015). In any case, a large body of evidence supports the use of pupil dilation as a measure of arousal/task engagement (Aston-Jones et al., 1994; 2007; Gilzenrat et al., 2010; Murphy et al., 2014; Unsworth & Robison, 2016; 2018).

The present study

The present study aimed to assess whether pupil dilation is a predictor of memory transfer effects in preparation, building on the idea that pupil dilation reflects arousal, alertness, and task engagement. Specifically, in the acquisition-transfer variable FP paradigm, we expected to see increased task engagement directly after instruction about the S1-FP contingency has taken place, which would gradually decrease as the experiment proceeded. This moment of increased task engagement would correspond to "the dip".

With regards to the pupil, our expectations were as follows: in the acquisition phase, participants are assumed to be mostly mind-wandering, so preparation is implicitly driven by

past trials via trace retrieval. Immediately after instruction, trace retrieval may be suppressed as participants behave according to explicit task goals, requiring voluntary control and therefore higher task engagement. This top-down control is associated with activity in the LC-NE system, which can be detected by an increase in tonic pupil dilation and larger phasic pupillary responses. Task engagement gradually decreases again in the following blocks as participants resume mind-wandering, reflected in a decreasing tonic pupil dilation.

Regarding the phasic pupil response, our study aimed to replicate and extend the finding that the pupil shows earlier and larger pupil dilation when participants expect a short FP (Akdoğan et al., 2016). In particular, we hypothesized that the pupil would demonstrate different transient responses to short vs. long cue S1s, indicating the effect of differential preparation.

In conclusion, transfer and cueing effects imply the involvement of long-term memory in temporal preparation. This is supported by the multiple trace theory of temporal preparation. To investigate the role of task engagement in temporal preparation, the present study applied pupillometry to an acquisition-transfer variable FP experiment. Pupil responses have been shown to reflect LC-NE activity and therefore provide a good measure of arousal/task engagement.

Method

Participants

Participants were recruited through the SONA participant pool of the psychology department at the University of Groningen, and completed the experiment in exchange for course credit. All participants gave informed consent at the start of the experiment. The study was deemed exempt from review based on criteria defined by the Ethics Committee Psychology. All procedures were conducted following the guidelines of the NETHICS code of conduct for experiments.

The experiment was successfully completed by 32 participants (18 female) between 17 and 26 years of age (M = 20.06, SD = 1.78). All participants had normal or corrected-to-normal eyesight.

Task and stimuli

Participants were seated in a dimly lit room with their head in a chin rest at 70 cm viewing distance from a 27-in. LCD monitor (Iiyama ProLite g2773hs-gb1) with 1024x768 resolution and 60-Hz refresh rate. The experiment was designed and run using OpenSesame version 3.2.8 (Mathôt et al., 2012).

Each trial started with space-bar triggered drift correction, followed by the presentation of a black fixation dot (a filled circle with an 0.20° radius and a 0.05° hole) against a gray background at the center of the screen for 500ms. This constituted the intertrial interval (ITI) and served to measure pre-stimulus baseline pupil size. Next, the S1 was presented, which was either a visual stimulus (a filled black circle with a radius of 1.16° at the center of the screen) or an auditory stimulus (a 440 Hz tone) for 145ms. Then, after a variable FP of either 700, 1100, or 1400ms, the S2 was presented. S2 was a black square (0.82° width and height), placed at 3.07° distance to the left or right of fixation with equal probability. Participants were instructed to press the *z* or *m* key when S2 appeared left or right of fixation, respectively, and to do so as quickly as possible while maintaining high accuracy. After the participant responded, S2 was immediately removed.

Design and procedure

The foreperiod (FP) on each trial was defined as the time between the onset of the warning stimulus (S1) and the target stimulus (S2) and was either 700, 1100, or 1400ms. FPs varied randomly within blocks according to one of three different distributions: uniform, exponential, or anti-exponential. In blocks with a uniform distribution, each FP duration has equal probability of occurrence. Exponential and anti-exponential distributions, on the other hand, are biased distributions: exponential distributions have predominantly trials with shorter FPs and anti-exponential distributions have predominantly trials with longer FPs. Concretely, this was implemented as a 4:2:1 proportion of 700/1100/1400ms FPs for the exponential distribution, and 1:2:4 for the anti-exponential distribution.

The experiment consisted of eight experimental blocks divided into three phases: the neutral phase (experimental block 1), the acquisition phase (blocks 2 to 5), and the transfer phase (blocks 6 to 8). In the acquisition phase, a visual S1 (circle) was paired with an exponential FP distribution and an auditory S1 (tone) was paired with an anti-exponential FP distribution; or vice versa, in accordance with counterbalancing. Through this pairing, participants were expected to learn the association between S1 type FP distribution. The neutral and transfer phases had a uniform FP distribution, thus S1 type had no association with expected FP duration. The transfer phase thus provided the opportunity to study the persistence of previously learned associations in the absence of the S1-FP relation.

At the end of the acquisition phase, in between the fifth and sixth experimental block, the experiment was paused and participants were asked about their awareness of the S1-FP association by means of a short questionnaire. Participants were first asked to describe if there was anything they noticed about the experiment in general. Next, participants were asked to guess the number of different FP durations that were present in the experiment. Finally, participants were asked a multiple choice question whether (1) Circles were more often followed by a short FP and tones by a long FP; (2) vice versa; (3) they did not know. After these three questions, participants were informed of the predictive value of S1 in the preceding blocks, and were told that this association would no longer hold in the final blocks of the experiment. After this 'intervention', the experiment continued with the transfer phase, starting with block 6.

Each experimental block consisted of 84 trials. Together with the practice block of 12 trials, the total experiment took participants approximately 60 minutes to complete.

Statistical analyses

Statistical analysis was conducted in three parts: behavioral analysis, tonic pupil response analysis, and phasic pupil response analysis. Practice trials, trials with incorrect responses, and outlier trials were discarded. For the behavioral analysis, outlier trials were defined as trials with an RT further than 3 SD away from the participant's mean RT. On average, 3.66% of trials were discarded per participant (SD = 0.97%). For the pupillary analysis, outlier trials were defined as trials with moment-to-moment changes in pupil size bigger than 5*MAD away from the median. On average, 24.45% of trials were discarded per participant (SD = 12.21%).

The behavioral analysis consisted of linear mixed effects model comparison as well as rolling regression, which is a way to analyse time course data on a moment-to-moment basis (see Kruijne et al., 2022). FP, S1 type and distribution were hypothesized to have an effect on RT. In our LMM analysis, we started with the complete model containing all main effects and interactions, and compared this with various reduced models by means of BIC scores. Additionally, rolling regression was applied to a linear model with four β-coefficients: the intercept, FP, S1 type, and their interaction. An 80-trial rolling window generated trial-wise estimates of each coefficient, resulting in four β-coefficient time courses. A baseline (determined by the first value) was subtracted from each time course, in order to quantify the development of the β-coefficient throughout the experiment. These β-time courses were then tested using cluster-based permutation t-testing, a non-parametric method of testing for significant changes in time-course data (Maris & Oostenveld, 2007). Clusters were deemed significant at $\alpha = 0.05$.

Tonic pupil size was also analyzed by means of cluster-based permutation testing. For each participant, a tonic pupil time course was calculated by taking the mean pupil size during the first 500ms of the ITI, to get one value per trial. Missing trials were interpolated using a sliding window of 3 trials. The overall tonic pupil time course was then calculated by averaging these values over all participants. This time course was split per block, so that block 6 (the first block after the intervention) could be compared to the other blocks. Then it was determined at which trials in the block there was a significant difference between the sixth block and the others.

Finally, analysis of the phasic pupil also took the form of cluster-based permutation testing. The first 700ms of each trial's pupillary data was split per phase, and compared per S1 type. We only took the first 700ms of each trial, because this segment was present in all trials, because 700ms is the length of the shortest FP. Pupil responses for long and short S1

categories were compared in order to determine clusters of trials at which these differences were significant.

Results

Questionnaire

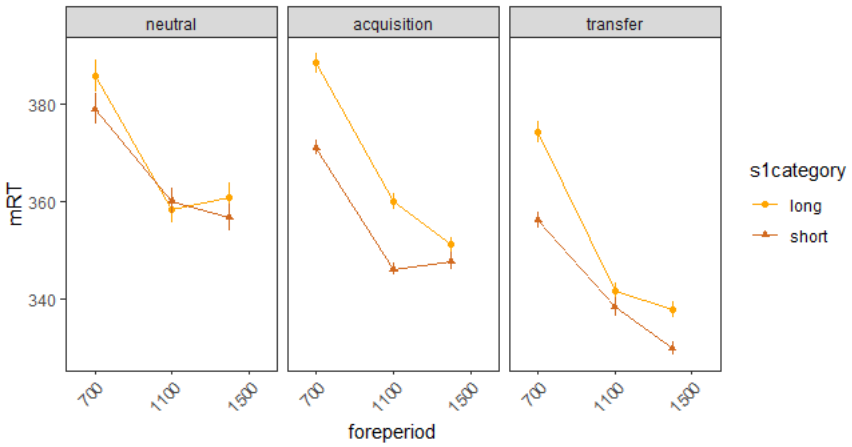
Just over half of the participants (53.13%) correctly answered that the experiment had three different FP durations. For the multiple-choice question, eight participants answered that they did not know which S1 type had been associated with which FP distribution. Of the remaining participants, 62.5% correctly identified the association between S1 and FP distribution. This proportion is similar to that found in earlier research (65%; see Kruijne et al., 2022), and does not seem to indicate awareness of the S1-FP contingency.

Only one participant's response to the open question indicated that they may have been aware of the S1-FP association: *"I noticed that when I hear the noise the square takes longer to appear than when I do not hear any noise."* They also answered the multiple choice question about the S1-FP association correctly. However, this subject guessed that there were only two possible FP durations. Therefore, they were not excluded from analysis.

Behavioral analysis

Figure 1 illustrates a number of effects visible throughout the experiment. The main effect that can be seen in each phase is the typical FP-RT curve, which reflects faster RTs for longer FPs. Furthermore, a gradual decrease in RT can be seen over the experiment. In the acquisition phase, the FP-RT curves for long and short S1s diverge, supporting the cueing effect. This effect persists in the transfer phase, consistent with the transfer effect.

Figure 1. *The RT-FP curve across phases.*



Effects on RT were analyzed by means of linear mixed-effect model comparisons. Two routes of analysis were taken: Satterthwaite's method (Satterthwaite, 1946) and BIC model comparison (Wagenmakers, 2007). Satterthwaite's method indicates that the full model (expressing RT as a function of FP, S1 type, Phase, and all possible interactions) is the best model. In this model, all main effects and interactions except for the three-way interaction are significant (see Table 1).

Table 1. *ANOVA on the full LMM of RT using Satterthwaite's method.*

	SS	*MS*	df_{num}	df_{den}	*F*	*p*
FP	3.19E-05	3.19E-05	1	38.9	139.901	1.89E-14 ***
S1 type	9.66E-06	9.66E-06	1	21036.1	42.399	7.61E-11 ***
phase	8.63E-05	4.32E-05	2	21036.1	189.431	2.20E-16 ***
FP * S1 type	9.96E-07	9.96E-07	1	21036.3	4.372	0.037 *
FP * phase	2.73E-06	1.36E-06	2	21036.4	5.985	0.003 **
S1 type * phase	1.64E-06	8.19E-07	2	21036.1	3.595	0.027 *
FP * S1 type * phase	6.65E-07	3.32E-07	2	21036.3	1.459	0.232

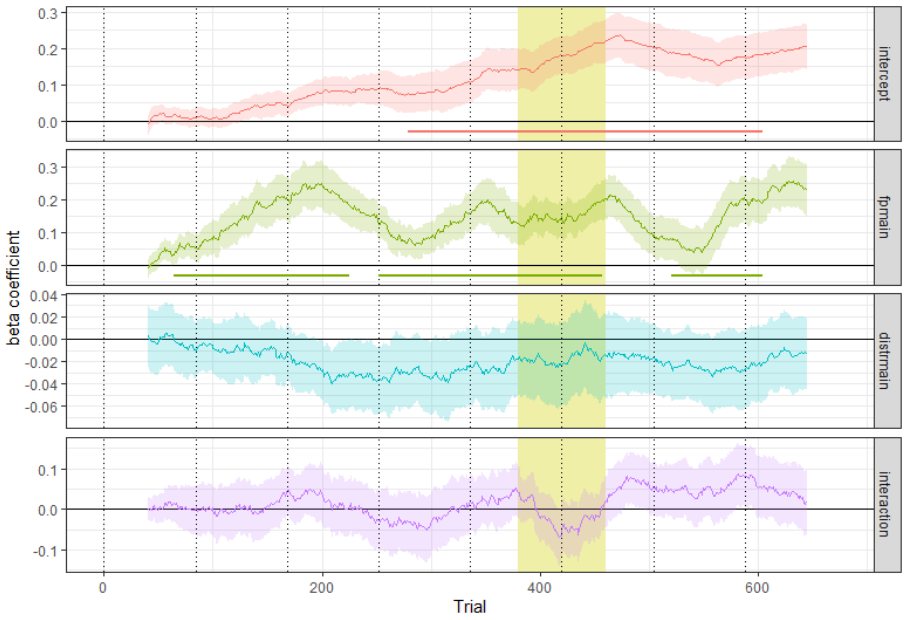
Note. *** indicates p < 0.001, ** indicates p < 0.01, * indicates p < 0.05.

A different model was found to be best when selected according to BIC. This model expressed RT as a function of FP, S1 type and their interaction, with an additional main effect of Phase. In this analysis, this model was preferred to the aforementioned full model (*ΔBIC* = 37.75; *BF* > 1000).

The critical finding from these two approaches is that both methods show that S1 type had a significant effect. It is unexpected that the interaction between S1 type and Phase was absent in the BIC-determined model. The BIC-determined model predicts an effect in the first block. Seeing as there is no theoretical reason why this would occur, given that no S1-FP associative learning has taken place at that point, it suggests that the study was simply underpowered to identify the Phase x S1 type interaction with BIC-based inference, which is quite conservative. Additionally, it should be taken into account when interpreting these results that the neutral phase consists of only one block, and therefore has relatively little observations compared to the other phases.

Rolling regression analysis resulted in Δβ time courses for all four model coefficients, showing the development of each coefficient per participant throughout the experiment. The mean time course of each coefficient is plotted in Figure 2. A cluster-based permutation test was applied to the coefficient time courses to determine at what time points the clusters significantly differ from baseline. Significant clusters are indicated in the figure as horizontal colored lines.

Figure 2. *Rolling regression results.*

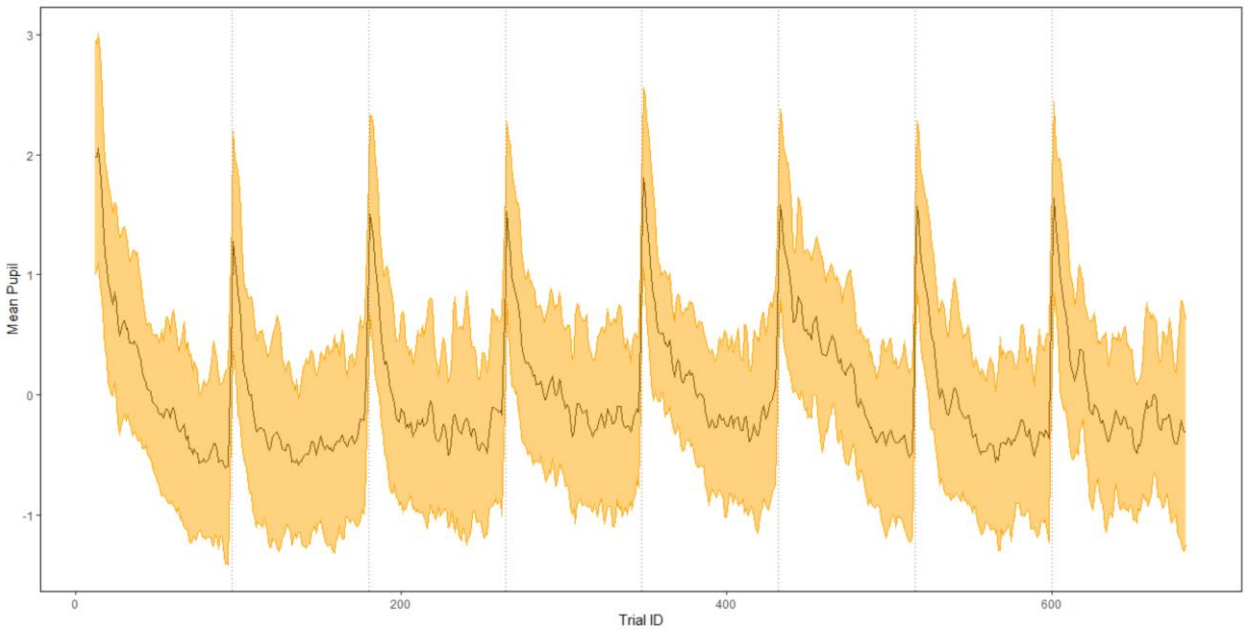


The gradual increase of the intercept coefficient reflects the faster average RT as the experiment proceeded. In particular, RT seemed to improve markedly from the sixth block onwards, so after participants are instructed about the S1-FP association. Furthermore, the fpmain (FP slope) coefficient fluctuated quite a bit throughout the experiment. A higher fpmain coefficient indicates a steeper RT-FP curve, whereas a lower fpmain coefficient indicates a flatter RT-FP curve. The slight decrease indicates that participants responded somewhat faster on trials with exponential distributions than on trials with anti-exponential distributions. Finally, the S1-FP interaction coefficient shows fluctuations as well. Notably, the interaction coefficient exhibits a dip around the start of the sixth block. This dip does not quantitatively match earlier findings (see Kruijne et al., 2022), but occurs at the crucial time, namely immediately after participants were informed about the S1-FP contingency. A similarly timed dip is also visible in the distrmain coefficient, albeit at a smaller scale than the interaction coefficient. This dip is especially notable considering that the S1 effect is one of only few effects found. The dip was seemingly not apparent in any of the other coefficients.

Tonic pupil response

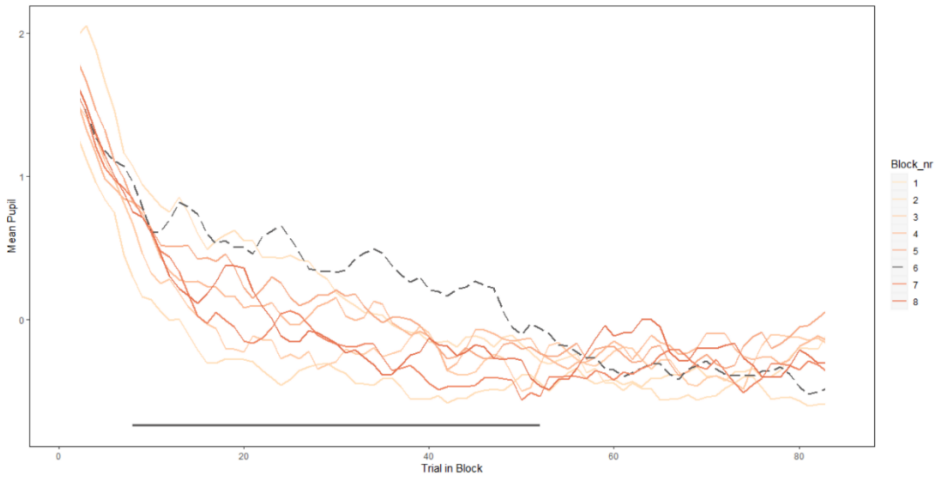
Figure 3 illustrates mean baseline pupil size throughout the experiment. Overall, tonic pupil size spikes at the beginning of each experimental block and quickly decreases to a baseline level that persists until the block ends. Notably, this decrease is less steep in block 6, compared to the other blocks, indicating that the pupil stayed dilated longer throughout the time course of this block.

Figure 3. *Time course of tonic pupil dilation throughout the experiment.*



To test whether the pupil indeed shows a larger dilation at the start of block 6 than other blocks, we performed a cluster-based permutation test with 1000 permutations, comparing block 6 with the other blocks (see Figure 4). Tonic pupil dilation in block 6 differed significantly at the start of the block, with a significant cluster spanning from trial 8 until 52 ($p = 0.001$; significant trials indicated by the gray horizontal line). Note that a significant cluster does not indicate that the effect exists in these trials only.

Figure 4. *Comparing tonic pupil dilation for each experimental block.*

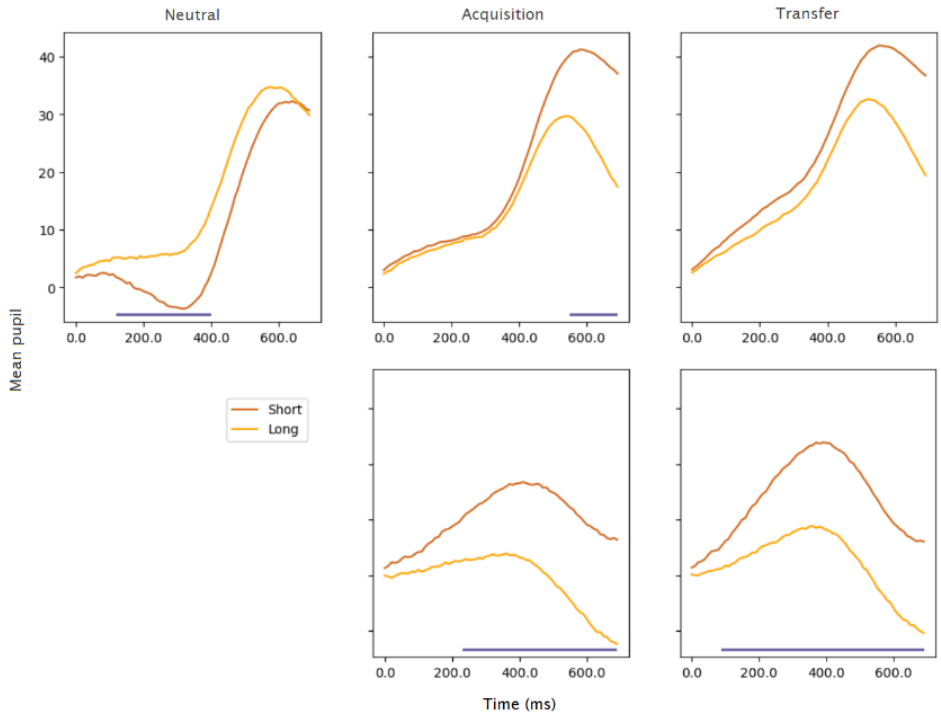


Phasic pupil response

To investigate the phasic pupil response throughout the experiment, we compared how pupil size developed in the 700ms following S1. This allows for comparison of all trials since the shortest FP is 700ms. We performed a cluster-based permutation test with 1000 permutations, resulting in a significant difference between the long cue and short cue S1 in the neutral phase ($p = 0.035$, 120-400ms) as well as an absence of large significant clusters in the acquisition and transfer phases. The finding that the long and short cue S1s significantly differed in the neutral phase could not have resulted from differential preparation as no S1-FP association was present yet. Furthermore, it was surprising that qualitatively, the difference between the two mean responses reversed in the acquisition and transfer phases without yielding significant results.

It is likely that the effect found in the neutral phase reflects the difference that auditory and visual S1s have on the pupil, given the imperfect counterbalancing of S1 type. Therefore, we decided to use the phasic pupil response of this phase as a baseline to isolate the differential preparation effect. By subtracting the baseline from the responses of the acquisition phase and transfer phase, we aimed to get a clearer picture of the preparation effect reflected by the pupil in isolation. Doing so resulted in significant clusters for acquisition phase ($p = 0.013$, 230-690ms) and transfer phase ($p = 0.004$, 90-690ms). This demonstrates that in the acquisition and transfer phases, participants' pupils dilated more when they were presented with a short-cue S1 than a long-cue S1. Figure 5 shows the three phases before subtracting the baseline, as well as the acquisition and transfer phases after subtracting the baseline. Significant clusters are indicated by the blue horizontal lines.

Figure 5. *Phasic (transient) pupil responses for long cue and short cue S1s.*



Note. The top row represents the original phasic responses, and the bottom row represents the phasic responses once the baseline (neutral phase) is subtracted.

Discussion and Conclusion

Previous research has demonstrated the involvement of long-term memory in temporal preparation (e.g. Kruijne et al., 2022; Los et al., 2017; 2022; Mattiesing et al., 2017), in accordance with the multiple trace theory of temporal preparation (Los et al., 2014; Salet et al., 2022). According to this theory, learned associations drive preparation through the automatic retrieval of memory traces. This leads to the question: are there factors that affect this automatic retrieval?

The present study investigated the link between memory transfer effects and arousal, by means of pupil dilation. We expected tonic pupil size to reflect the changes in task engagement throughout the experiment; in particular, we expected tonic pupil size to peak at the beginning of the experiment and after instruction about the S1-FP association had taken place.

Contrary to expectations, tonic pupil size did not decrease through the experiment as a whole. However, we did find that tonic pupil size peaked at the beginning of each block, decreasing rapidly in the following trials. The exception was the development of tonic pupil size in block 6 (post-intervention), which showed a significantly slower decrease compared to the other blocks. Previous behavioral studies (e.g. Kruijne et al., 2022) suggest that the informed change from the acquisition to the transfer phase evokes a momentary suppression of automatic trace retrieval, accompanied or caused by a brief burst of arousal. The larger pupil size we found in block 6 shows that there might indeed be an increase in arousal following the S1-FP association instruction.

Furthermore, we found that the phasic (within-FP) pupil response to short versus long S1s significantly differed from baseline in the acquisition and transfer trials, suggesting that learning and persistence of the S1-FP association can be detected via the pupil dilation measure. Since participants seemed generally unaware of this association, it is all the more notable that the resulting associative guidance was detected in fluctuations of the pupil response.

Some limitations must be noted. First, despite the effects of preparation reflected in the pupil, we did not find a one-to-one link between pupil size and behavior. We did find moments in which the two vary together. The pupil reflects more than arousal, and disentangling the various components present in the pupil response could provide more insights in the exact relationship between preparation and arousal.

Furthermore, the shortest FP used in this experiment (700ms) is quite long compared to earlier studies, but it was chosen to capture the transient pupillary response. Based on the results we found, one could argue that 700ms was still too short to provide a complete picture of the phasic response. On the other hand, the shortest FP being 700ms is a likely cause for why the memory effect came to expression as a main effect, more so than an interaction (S1 type x FP) effect. This is because, taking into account the shape of the FP-RT curve, larger FPs show a smaller interaction effect for the different FP distributions. Shorter FPs may more accurately capture the interaction effect. Future studies should consider this trade-off when choosing FP durations.

Additionally, the inter-trial interval (ITI) had a fixed duration of 500 ms on each trial. The steep slopes visible at the beginning of the phasic pupil response in the FPs of the acquisition and transfer phase (see figure 5) indicate that participants already show markers of preparing – namely a transient increase in pupil size – during the ITI, as they have learned that S1 appears exactly 500 ms after drift correction. This may explain why no direct link

between behavioral effects and the baseline pupil was found. Future studies should consider jittering the ITI to combat this issue.

Finally, a considerable limitation of the current study was its low power. While our sample size of 32 participants was comparable to earlier studies (e.g. Los et al., 2021; Mattiesing et al., 2017), participants completed relatively few trials. Future studies should either extend the duration of the experiment or use shorter ITIs. Furthermore, the magnitudes of the studied preparation effects are unknown. This should therefore be taken into account when interpreting our results.

Future research could incorporate different experimental manipulations. One example involves distinguishing between rewarding and non-rewarding trials or blocks in order to influence motivation. Using the acquisition transfer design, one could expect less memory-guided preparation following reward, given the increased arousal induced by the reward, whereas non-rewarding blocks result in low motivation, and therefore show less memory effects. This design could test the relationship between arousal and memory effects more directly.

Future studies could also incorporate alternative measurements of arousal and/or task engagement. For example, task engagement could be measured throughout the experiment using thought probes (Kane et al., 2017; Weinstein, 2018). This method has been successfully applied to pupillometry research by Unsworth and Robison (2016), who were able to distinguish on-task states from off-task states using tonic pupil measures. Another option is to use EEG, in which case neural markers of cognitive control could be linked to memory effects of preparation (Miller & Cohen, 2001; Zavala et al., 2018). This could help determine more precisely how arousal fluctuates during the experiment, and how this relates to preparation.

In conclusion, the multiple trace theory of temporal preparation provides a solid framework to incorporate the role of task engagement in the study of preparation effects. Our results suggest that there exists a link between arousal and memory effects in preparation, although is not a direct link. This may have to do with the use of pupil dilation as an indirect measure of arousal, given that pupil dilation reflects more than just arousal. Future studies can further investigate this topic using the same measure, but other methods, e.g. EEG, fMRI, skin conductance, or any combination of methods, may generate more and different insights. In this way, future research can extend and verify the present findings.

References

Akdoğan, B., Balcı, F., & van Rijn, H. (2016). Temporal expectation indexed by pupillary response. *Timing & Time Perception*, *4*(4), 354-370.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.*, *28*, 403-450.

Aston-Jones, G., Iba, M., Clayton, E., Rajkowski, J., & Cohen, J. (2007). The locus coeruleus and regulation of behavioral flexibility and attention: clinical implications. *Brain Norepinephrine: Neurobiology and Therapeutics.*

Aston-Jones, G., Rajkowski, J., Kubiak, P., & Alexinsky, T. (1994). Locus coeruleus neurons in monkey are selectively activated by attended cues in a vigilance task. *Journal of Neuroscience*, *14*(7), 4467-4480.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin*, *91*(2), 276.

Berridge, C. W., & Waterhouse, B. D. (2003). The locus coeruleus–noradrenergic system: modulation of behavioral state and state-dependent cognitive processes. *Brain research reviews*, *42*(1), 33-84.

Bradshaw, J. L. (1968). Pupillary changes and reaction time with varied stimulus uncertainty. *Psychonomic Science*, *13*(2), 69-70.

Chatham, C. H., Frank, M. J., & Munakata, Y. (2009). Pupillometric and behavioral markers of a developmental shift in the temporal dynamics of cognitive control. *Proceedings of the National Academy of Sciences*, *106*(14), 5529-5533.

Cohen, J. D., Aston-Jones, G., & Gilzenrat, M. S. (2004). A systems-level perspective on attention and cognitive control: Guided activation, adaptive gating, conflict monitoring, and exploitation versus exploration.

Crowe, E. M., & Kent, C. (2019). Evidence for short-term, but not long-term, transfer effects in the temporal preparation of auditory stimuli. *Quarterly Journal of Experimental Psychology*, *72*(11), 2672-2679.

Crowe, E. M., Los, S. A., Schindler, L., & Kent, C. (2021). Transfer effects in auditory temporal preparation occur using an unfilled but not filled foreperiod. *Quarterly Journal of Experimental Psychology*, *74*(8), 1432-1438.

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(2), 252-269.

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science*, *132*(3423), 349-350.

Janssen, P., & Shadlen, M. N. (2005). A representation of the hazard rate of elapsed time in macaque area LIP. *Nature neuroscience*, *8*(2), 234-241.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154*(3756), 1583-1585.

Kane, M. J., Gross, G. M., Chun, C. A., Smeekens, B. A., Meier, M. E., Silvia, P. J., & Kwapil, T. R. (2017). For whom the mind wanders, and when, varies across laboratory and daily-life settings. *Psychological science*, *28*(9), 1271-1289.

Knapen, T., de Gee, J. W., Brascamp, J., Nuiten, S., Hoppenbrouwers, S., & Theeuwes, J. (2016). Cognitive and ocular factors jointly determine pupil responses under equiluminance. *PloS one*, *11*(5), e0155574.

Koss, M. C. (1986). Pupillary dilation as an index of central nervous system α2-adrenoceptor activation. *Journal of pharmacological methods*, *15*(1), 1-19.

Kruijne, W., Galli, R. M., & Los, S. A. (2022). Implicitly learning when to be ready: From instances to categories. *Psychonomic bulletin & review, 29*(2), 552-562.

Laeng, B., Ørbo, M., Holmlund, T., & Miozzo, M. (2011). Pupillary stroop effects. *Cognitive processing*, *12*, 13-21.

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious?. *Perspectives on psychological science*, *7*(1), 18-27.

Los, S. A. (2013). The role of response inhibition in temporal preparation: Evidence from a go/no-go task. *Cognition*, *129*(2), 328-344.

Los, S. A., & Agter, F. (2005). Reweighting sequential effects across different distributions of foreperiods: Segregating elementary contributions to nonspecific preparation. *Perception & Psychophysics*, *67*, 1161-1170.

Los, S. A., Knol, D. L., & Boers, R. M. (2001). The foreperiod effect revisited: Conditioning as a basis for nonspecific preparation. *Acta Psychologica*, *106*(1-2), 121-145.

Los, S. A., Kruijne, W., & Meeter, M. (2014). Outlines of a multiple trace theory of temporal preparation. *Frontiers in Psychology, 5*, 1058.

Los, S. A., Kruijne, W., & Meeter, M. (2017). Hazard versus history: Temporal preparation is driven by past experience. *Journal of Experimental Psychology: Human Perception and Performance, 43*(1), 78.

Los, S. A., Nieuwenstein, J., Bouharab, A., Stephens, D. J., Meeter, M., & Kruijne, W. (2021). The warning stimulus as retrieval cue: The role of associative memory in temporal preparation. *Cognitive Psychology, 125*, 101378.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of neuroscience methods*, *164*(1), 177-190.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314-324.

Mattiesing, R. M., Kruijne, W., Meeter, M., & Los, S. A. (2017). Timing a week later: The role of long-term memory in temporal preparation. *Psychonomic Bulletin & Review, 24*(6), 1900-1905.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, *24*(1), 167-202.

Murphy, P. R., O'connell, R. G., O'sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human brain mapping, 35*(8), 4140-4154.

Näätänen, R. (1971). Non-aging fore-periods and simple reaction time. *Acta psychologica*, *35*(4), 316-327.

Niemi, P., & Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological bulletin, 89*(1), 133.

Nieuwenhuis, S., De Geus, E. J., & Aston-Jones, G. (2011). The anatomical and functional relationship between the P3 and autonomic components of the orienting response. *Psychophysiology*, *48*(2), 162-175.

Nobre, A. C., Correa, A., & Coull, J. T. (2007). The hazards of time. *Current opinion in neurobiology*, *17*(4), 465-470.

Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology*, *47*(3), 560-569.

Privitera, C. M., Renninger, L. W., Carney, T., Klein, S., & Aguilar, M. (2008). The pupil dilation response to visual detection. *Human Vision and Electronic Imaging XIII, 6806*, 290-300.

Salet, J. M., Kruijne, W., van Rijn, H., Los, S. A., & Meeter, M. (2022). FMTP: A unifying computational framework of temporal preparation across time scales. *Psychological Review*, *129*(5), 911.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics bulletin*, *2*(6), 110-114.

Shalev, N., & Nobre, A. C. (2022). Eyes wide open: Regulation of arousal by temporal expectations. *Cognition*, *224*, 105062.

Tressoldi, P. E., Martinelli, M., Semenzato, L., & Cappato, S. (2011). Let your eyes predict: prediction accuracy of pupillary responses to random alerting and neutral sounds. *Sage Open*, *1*(2), 2158244011420451.

Trillenberg, P., Verleger, R., Wascher, E., Wauschkuhn, B., & Wessel, K. (2000). CNV and temporal uncertainty with ‘ageing’and ‘non-ageing’S1–S2 intervals. *Clinical neurophysiology*, *111*(7), 1216-1226.

Unsworth, N., & Robison, M. K. (2016). Pupillary correlates of lapses of sustained attention. *Cognitive, Affective, & Behavioral Neuroscience*, *16*, 601-615.

Unsworth, N., & Robison, M. K. (2018). Tracking arousal state and mind wandering with pupillometry. *Cognitive, Affective, & Behavioral Neuroscience*, *18*, 638-664.

Usher, M., Cohen, J. D., Servan-Schreiber, D., Rajkowski, J., & Aston-Jones, G. (1999). The role of locus coeruleus in the regulation of cognitive performance. *Science*, *283*(5401), 549-554.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic bulletin & review*, *14*(5), 779-804.

Wang, C. A., & Munoz, D. P. (2015). A circuit for pupil orienting responses: implications for cognitive modulation of pupil size. *Current opinion in neurobiology*, *33*, 134-140.

Weinstein, Y. (2018). Mind-wandering, how do I measure thee with probes? Let me count the ways. *Behavior research methods*, *50*, 642-661.

Woodrow, H. (1914). The measurement of attention. *The psychological monographs, 17*(5), i.

Zahn, T. P., & Rosenthal, D. (1966). Simple reaction time as a function of the relative frequency of the preparatory interval. *Journal of Experimental Psychology*, *72*(1), 15.

Zavala, B., Jang, A., Trotta, M., Lungu, C. I., Brown, P., & Zaghloul, K. A. (2018). Cognitive control involves theta power within trials and beta power across trials in the prefrontal-subthalamic network. *Brain*, *141*(12), 3361-3376.