



Time-of-day effects on eyewitness reports in morning and evening types

Sergii Yaremenko^{a,b} , Melanie Sauerland^a  and Lorraine Hope^b 

^aThe Department of Clinical Psychological Science, Maastricht University, Maastricht, the Netherlands; ^bDepartment of Psychology, University of Portsmouth, Portsmouth, UK

Our performance varies throughout the day as a function of alignment with our circadian rhythms. The current experiment tested whether similar performance patterns can be observed in eyewitness memory performance. One-hundred-and-three morning-type and evening-type participants watched a stimulus event, provided a free report and answered cued questions in the morning and the evening hours, respectively. We expected eyewitness reports to be more detailed and more accurate at participants' circadian peaks than at circadian troughs. Contrary to our predictions, time of testing did not significantly affect quantity and accuracy of eyewitness statements. Future studies might investigate whether matching chronotype with time of day would be beneficial when encoding or retrieval conditions are suboptimal or when eyewitnesses are vulnerable.

Keywords: body clock; chronotype; circadian rhythm; cued recall; eyewitness memory; free recall; synchrony effect; time of day.

Introduction

Daily variations in our physiology, behaviour and cognition are inherent in our daily lives. These changes do not happen chaotically: a central 'pacemaker' of our body, known as *the circadian clock*, ensures that the numerous systems in our body function in synchrony with each other and with the external environment (Halberg et al., 2003). Located in the suprachiasmatic nucleus of the hypothalamus, the circadian clock maintains 24-hour cycles in many aspects of our physiology and behaviour (Czeisler & Gooley, 2007; Fisk et al., 2018).

To a large extent, it is by virtue of the circadian clock that we experience peaks and dips in alertness and performance throughout the day. Some hours of the day are more optimal for engaging in cognitively demanding activities than others. This daily variation in

cognitive performance is known as *the synchrony effect*, referring to the fact that the timing of our activities can be synchronised with our internal pacemaker to varying degrees. Performance is better whenever the timing of the task is congruent with the circadian phase, whereas performance is lower when the timing of the task is incongruent with our internal body clock (Correa et al., 2020; Schmidt et al., 2007).

The circadian clock can be 'set' slightly differently in some individuals compared to others. The so-called morning types, also known as '*larks*' wake up and go to sleep earlier than others and prefer to be more active in the morning hours. Evening types, or '*owls*', on the contrary, prefer to wake up and go to sleep later, and are at their best later in the evening (Horne & Östberg, 1976). This time-of-day preference is known as *the circadian*

typology, or *chronotype* (Adan et al., 2012). A relatively large proportion of the adult population (about 40–50%) belong to either a morning- or an evening-chronotype group (Paine et al., 2006; Roenneberg et al., 2007).

Circadian performance patterns occur in many aspects of our mental life, including attention, working memory, inhibition, engagement in risky decision-making and even proneness to rely on heuristics and biases (for a review, see Schmidt et al., 2007). Episodic memory is also subject to the synchrony effect: recall and recognition performance is better at circadian peaks as opposed to circadian troughs (e.g. May et al., 2005; Puttaert et al., 2019; Yang et al., 2007). Specifically, free recall performance can vary as a function of alignment of time of testing and an individual's optimal performance timings. In one experiment, morning-type and evening-type participants studied prose passages and wrote down what they remembered in their own words. Testing sessions were scheduled at 9 am, 2 pm, or 8 pm. Morning-type participants produced significantly more idea units (defined as ideas or places where a reader might pause; Brown & Smiley, 1977) from the passages at 9 am than from those in the later sessions. Although the effect for evening-type participants was non-significant, descriptives showed a tendency in the expected direction (Petros et al., 1990).

In another study (May et al., 2005), younger morning-types and older evening-types encoded target words either in the morning (8 am and 9 am) or in the early evening (between 5 pm and 6 pm). When presented with a surprise stem completion task, participants in both chronotype groups showed the standard synchrony effect pattern – that is, morning-type participants recalled more studied words in the morning, whereas evening types completed more stems correctly in the evening. Similar results in other experiments confirmed that cued recall (Puttaert et al., 2019; Yang et al., 2007) and recognition of

verbal stimuli (May et al., 1993) follow synchrony effect patterns in performance.

Research into the circadian arousal patterns and their effect on memory can also have important implications outside the psychological laboratory. Adolescent education is one area where adjusting activity patterns with our internal body rhythms can aid efficient learning and high intellectual performance (Goldstein et al., 2007). Eyewitness testimony represents another example of an applied domain that relies on the functioning of episodic memory. Can witnesses remember events differently depending on the hour of the day when the event occurred? Can we help eyewitnesses provide more reliable testimony by aligning the time of interview with their natural performance patterns? The recall and recognition literature suggests that time-of-day optimality can produce variations in performance with a small to medium effect size (e.g. May et al., 2005; May et al., 1993; Puttaert et al., 2019; Ryan et al., 2002) comparable, for instance, to the sizes of effects of high relevance to the eyewitness memory field, such as biased identification instructions or weapon focus effect (Stebly, 1992, 1997).

Although there are a number of studies on the effects of sleep on eyewitness memory (Blagrove, 1996; Frenda et al., 2014; Morgan et al., 2019), to the best of our knowledge, only one published study tested the possibility of circadian variations in eyewitness memory performance (Diges et al., 1992). Mock eyewitnesses with morning or evening chronotype came to the lab at 10 am or at 8 pm. They encoded a stimulus film depicting a traffic accident and provided free narratives and answers to cued questions about the witnessed event. Contrary to predictions of the synchrony effect literature, results showed an overall pattern of better performance in the morning than in the evening regardless of chronotype. Unfortunately, the paper does not report participants' mean scores on the Morningness–Eveningness Questionnaire (MEQ). We therefore cannot assess whether

this tendency was related to possible differences in strength of time-of-day preference in each chronotype. More specifically, morning types might have shown a stronger preference for the morning, compared to evening types' preference for the late hours. Furthermore, the sample size was very small with 10 participants per experimental condition, and the method of coding and exclusion criteria were unclear.

In light of such a limited literature, the current experiment sought to further investigate the role of time of testing on the accuracy of eyewitness statements. Morning- and evening-type participants encoded two stimulus events and provided testimony at the time of day that once matched and once mismatched their peak circadian arousal periods. Based on previous research into the synchrony effect in memory performance, we expected participants to provide more detailed and more accurate free narratives and answers to cued questions at their optimal than at their non-optimal hours of the day.

Method

Participants

To determine the required sample size, we conducted a priori power analysis for a two-tailed paired-samples *t* test with G*Power v3.1 (Faul et al., 2009; Faul et al., 2007). Based on the data reported in previous studies on circadian effects in memory performance (May et al., 1993; Petros et al., 1990), we expected a small to medium effect size. Hence, we used an effect size *d* of 0.36, $\alpha = .05$ and a power of .95, resulting in a required sample size of 103.

To achieve the planned sample size, we recruited participants who self-identified as either morning or evening types using advertisements on a university notice board and by actively handing out flyers. Two-hundred-and-three individuals expressed their interest in participation and were pre-screened for their circadian typology using the short form of the

MEQ (rMEQ; Adan & Almirall, 1991). One-hundred-and-three pre-screened participants whose rMEQ score was ≤ 12 (evening types) or ≥ 17 (morning types) were invited to participate (15 male, 87 female, 1 unspecified; age 18–58 years, $M = 22.6$, $Mdn = 22$). The sample consisted of university students ($n = 98$) and members of the general public ($n = 4$). About half of the sample were evening-type (54.3%, $n = 56$, $M_{rMEQ} = 9.82$, $SD_{rMEQ} = 1.88$, age 19–29 years, $M = 22.2$, $Mdn = 22$) and morning-type participants (45.6%, $n = 47$, $M_{rMEQ} = 18.6$, $SD_{rMEQ} = 1.50$, age 18–58 years, $M = 23.4$, $Mdn = 21$). Figure 1 shows the density plot for distribution of rMEQ scores. Participants were native Dutch ($n = 56$), German ($n = 26$) or English ($n = 21$) speakers. All the recall instructions and questions were presented in the participants' native language, and participants were given a choice to provide responses in their native language. One participant showed inconsistent responses in the rMEQ compared to the full version of the questionnaire, and two participants did not attend the second testing session. We excluded data from these participants from the analyses.

Design

The experiment used a repeated measures design, with time-of-day optimality serving as predictor. Each participant was tested at both their optimal and the non-optimal time of day. The order of optimality conditions (optimal–non-optimal versus non-optimal–optimal) was counterbalanced to control for potential learning effects. In each of the sessions, participants encoded one of the two stimulus events. The order of presentation of stimulus films (Film 1–Film 2 versus Film 2–Film 1) was counterbalanced across optimality conditions. The number of correct details provided and the accuracy of eyewitness statements (correct items divided by incorrect items) served as dependent variables.

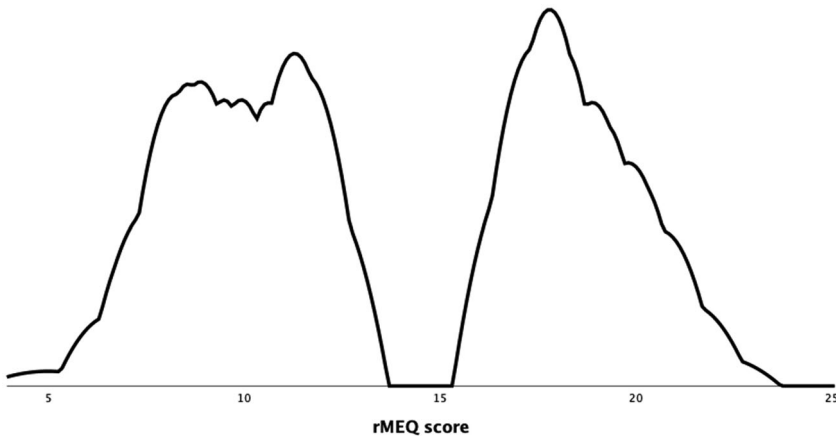


Figure 1. Density plot for distribution of rMEQ (reduced version of the Morningness–Eveningness Questionnaire) scores in participants with evening and morning chronotype in the final sample.

Materials

Morningness–eveningness scales

We used the rMEQ (Adan & Almirall, 1991) to classify participants into morning- and evening-type categories. The rMEQ consists of five items drawn from the original full 19-item Morningness-Eveningness Questionnaire (MEQ; Horne & Östberg, 1976). Both the MEQ and its reduced version are commonly used to assess individual differences in diurnal preferences with good external and construct validity (Adan & Almirall, 1991; Di Milia et al., 2013). The use of the shorter scale allowed us to distract participants’ attention from the main hypothesis by combining the rMEQ items with filler questions about eating habits (e.g. ‘When you get up in the middle of the night, how often do you snack?’). The rMEQ score ranges between 4 and 25, with high scores referring to stronger morningness preference. We adopted cut-offs of ≤ 12 for evening types and ≥ 17 for morning types, as opposed to those originally suggested of ≤ 11 for evening types and ≥ 18 for morning types (Adan & Almirall, 1991). Guided by the debate around the arbitrariness of the cut-offs suggested by the authors of questionnaires measuring diurnal preferences (Caci et al., 2009), we adopted more lenient cut-offs to increase the generalisability of our findings.¹

Additionally, we administered the full MEQ post hoc following all the experimental manipulations as an extra validation of our classification of participants into chronotype groups. To establish test–retest reliability, we extracted participants’ responses to the five rMEQ items from a full version of the MEQ questionnaire that was administered at the end of the experiment. The results showed excellent test–retest reliability, $r(98) = .92, p < .005$.

Stimulus films

Two different stimulus films depicting the theft of a wallet were used. The films differed in the details of the event, the environment and the actors. Film 1 (adapted from Sauerland et al., 2014) depicts a theft taking place at a bar. Four amateur actors (2 male, 2 female, 22–58 years old) appear in Film 1. In stimulus Film 2 (adapted from Brackmann et al., 2019) the theft occurs in a university communal area. Three amateur actors (2 male, 1 female, 21–26 years old) appear in Film 2.

Procedure

Data collection was carried out between March and June. All participants provided informed consent to participate in the

experiment. Participants were informed that the experiment focused on the way eating habits and caffeine consumption affect memory performance among early birds and night owls. They were instructed to exclude alcohol or caffeine-containing products and sleep a minimum of 6 hours prior to testing. Each participant attended the laboratory on two separate occasions, once in the morning (between 8 am and 10 am) and once in the evening (between 7 pm and 9 pm). The second session was always scheduled at least 36 hours after the first session in order to avoid possible fatigue effects. The protocol for the two sessions was analogous, except where specifically indicated.

First, participants encoded one of the two stimulus films. We asked participants to watch the film carefully and pay attention to every detail and informed them that they would be asked to act as eyewitnesses. Next, participants provided a free narrative of what they remembered about the event. Specifically, they were asked to report all the details they remembered about the incident, including the sequence of actions and events. They were also asked to describe the appearance of the people involved. Participants were asked to make their report as complete and accurate as possible and were discouraged from guessing. They had unlimited time to provide the free narratives.

After providing free narratives, participants went on to answer blocks of cued questions about the event and the people involved. First, they were presented with nine cued recall questions about the event (e.g. ‘Describe any interactions the thief/thieves had with the other people in the film’). Next, we presented participants with a schematic of the crime scene with people involved in the incident represented as silhouettes and informed them that they would be asked to answer questions about each person they saw in the film. Participants were presented with three (Film 1) or four (Film 2) blocks of cued questions about the appearance of each of the persons involved in

the incident, including their age, height, build, clothing and so on. Blocks of questions about each of the persons involved in the event were presented separately in the following order: thief, victim, Bystander 1, and (for Film 1) Bystander 2. For each of the blocks, we cued participants with the schematic of the crime scene, where the silhouette of the respective person was highlighted (i.e. the block of questions about the bystander was preceded with a schematic of the crime scene with highlighted silhouette of the bystander). The [Appendix](#) shows a complete list of cued questions.

After providing answers to cued questions, participants filled in either a sleep quality questionnaire (Session 1) or demographic questionnaires (Session 2), followed by a visual version of the Deese–Roediger–McDermott (DRM) paradigm (DRM; Moritz et al., 2006; Roediger & McDermott, 1995) task (about 15 min) and three or four eyewitness identification tasks (about 5 min; both sessions). Results from the identification tasks and the DRM paradigm are reported elsewhere (Yaremenko et al., 2021a, 2021b). At the end of Session 2, participants filled in the full MEQ and received either participation credit of gift vouchers worth 27.50 euros. The debriefing took place via email after data collection was terminated.

Recall coding

Following A. M. Wright and Holliday (2007), we developed a scoring template for both stimulus events. Each information unit from the events was categorised as an action (A), person (P), or object/setting (O/S) detail. For example, a stimulus film sequence of a perpetrator taking a wallet from the table was coded as ‘The guy (1-P) took (1-A) the wallet (1-O/S) from the table (1-O/S)’. The scoring templates contained 314 details for Film 1 and 298 details for Film 2.

Details reported during free recall and answers to cued questions were coded against the template for accuracy. A detail was coded as correct if it was present in the stimulus

Table 1. Quantity and accuracy of free and cued recall at optimal and non-optimal time of day.

	Time of day			
	Optimal		Non-optimal	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Free recall				
Number correct details	72.04	18.83	70.28	19.04
Number incorrect details	3.08	2.88	2.79	2.48
Total number details	75.98	19.91	73.75	19.86
Accuracy	.95	.04	.95	.04
Cued recall				
Number correct details	56.45	11.03	55.60	10.65
Number incorrect details	8.35	3.28	8.54	3.81
Total number details	64.98	11.54	64.37	11.64
Accuracy	.87	.05	.86	.05
Total recall				
Number correct details	101.43	17.87	99.51	16.44
Number incorrect details	10.27	4.40	10.20	4.49
Total number details	112.46	19.12	110.39	17.11
Accuracy	.90	.04	.90	.05

event and described correctly. Details that were present in the stimulus event but described incorrectly were coded as incorrect. Details described by participants that were not present in the stimulus event were coded as fabricated. Subjective responses (e.g. ‘The girl looked sad’) were excluded from analyses.

To establish inter-coder reliability, for each of the three languages, seven randomly selected statements from both stimulus films were coded by two independent scorers. Interrater reliability ranged from Cohen’s $\kappa = .70$ to $.87$, $ps < .001$, indicating substantial to almost perfect strength of agreement (Sim & Wright, 2005). When computing total statement accuracy, each detail provided by participants was counted once across free and cued recall.

Results

We ran paired-samples *t* tests to determine whether there was a statistically significant mean difference in the number of details and accuracy of statements provided at optimal

and non-optimal time of day. Additionally, we computed JZS Bayes factors (BFs) as a measure of a degree to which our data favour the null (BF_{01}) compared with the alternative (BF_{10}) hypothesis (Jeffreys, 1961). We applied default Cauchy’s prior with scaling factor of 0.707.

Table 1 shows that participants provided a large number of details both in the free recall and the cued questioning phase. Accuracy was generally high, but significantly higher in free than cued recall, $t(99) = 21.62$, $p < .001$, $d = 2.16$. Contrary to our hypothesis, there was no statistically significant effect of time-of-day optimality on the number of correct details reported during free recall, $t(99) = 1.13$, $p = .262$, $d = 0.11$, $BF_{01} = 2.82$, or in answers to cued questions, $t(99) = 0.62$, $p = .538$, $d = 0.06$, $BF_{01} = 5.15$. The Bayes factors suggest weak to moderate evidence in support of the null hypothesis (Lee & Wagenmakers, 2013). Testing optimality also did not have an effect on accuracy in free recall, $t(99) = -0.50$, $p = .618$, $d = -0.05$, $BF_{01} = 12.84$, or in answers to cued questions, $t(99) = 0.46$,

$p = .647$, $d = 0.05$, $BF_{01} = 6.05$. The Bayes factors indicate substantial to strong evidence in support of the null hypothesis (Lee & Wagenmakers, 2013). Indeed, accuracy in optimal and non-optimal sessions were strikingly similar. Figure 2 depicts the likelihood of observing these data if the true effect size is small, medium, or large, compared to the null hypothesis of no effect of testing optimality on recall quantity and accuracy. The charts show that if the true effect is small sized, the data are 0.7 to 1.8 times more likely to be observed than if the effect was zero.

Time-of-day optimality may have been more strongly pronounced in participants with more extreme morningness and eveningness. In order to explore this possibility, we ran Pearson's product-moment correlations to assess the relationship between participants' rMEQ scores and their recall performance. No significant correlations were observed in free recall statements, $ps > .05$. In cued recall statements, rMEQ score in morning types was positively associated with the number of correct details, $r(44) = .397$, $p = .006$, overall number of details, $r(44) = .309$, $p = .036$, and recall accuracy, $r(44) = .302$, $p = .041$, at a non-optimal time of day. No such association was present in optimal sessions, $ps > .05$. That is, when tested non-optimally, participants with stronger morningness preference were more likely to report an overall larger number of details and be more accurate. This finding was contrary to our hypotheses.

Discussion

In a two-session experiment, we tested eyewitness performance in morning- and evening-type participants. We expected enhanced quantity and accuracy of eyewitness reports during participants' circadian peaks compared to their circadian troughs. This hypothesis was based on previous research that consistently showed such a pattern of results for various recall and recognition tasks (May et al., 1993; Puttaert et al., 2019; Ryan et al., 2002; Yang

et al., 2007). Contrary to our predictions, time of testing had no effect on eyewitness performance in morning- and evening-type participants.

What might be possible explanations for these unexpected findings? Previous research on the synchrony effect mainly relied on verbal learning paradigms. Our experiment, on the other hand, used the eyewitness memory paradigm. There are some notable differences between the two. For example, we informed participants before presenting the stimulus event that they would be asked to serve as eyewitnesses. This situation mimics real-life incidents that are easily identifiable as crimes from the outset. Unlike the instructions used in previous synchrony effect studies, our procedure could activate participants' schemata of the role of eyewitnesses in criminal investigations. Processing the to-be-encoded stimuli in relation to pre-existing schemata generally has beneficial effects on encoding (e.g. Hastie & Kumar, 1979; see van Kesteren et al., 2012, for a review).

Informing participants that they would serve as eyewitnesses might also have encouraged a more elaborate processing of the presented stimuli than of the verbal stimuli used in previous studies. Based on common knowledge about eyewitness testimony, our participants could anticipate that they would be expected to report specific details from the to-be-presented stimuli, such as the appearance of the perpetrator, modus operandi, the sequence of events and so on. Such prior knowledge of the nature of the memory test may have affected our participants' encoding strategies, encouraging efficient distribution of attentional resources to enhance task-specific retrieval accuracy.

The identification data from this experiment support the idea that participants may have been aware of the potential detrimental effect of non-optimal testing and taken this information into account (see Experiment 1; Yaremenko et al., 2021a). Specifically, participants who made a selection from line-ups with

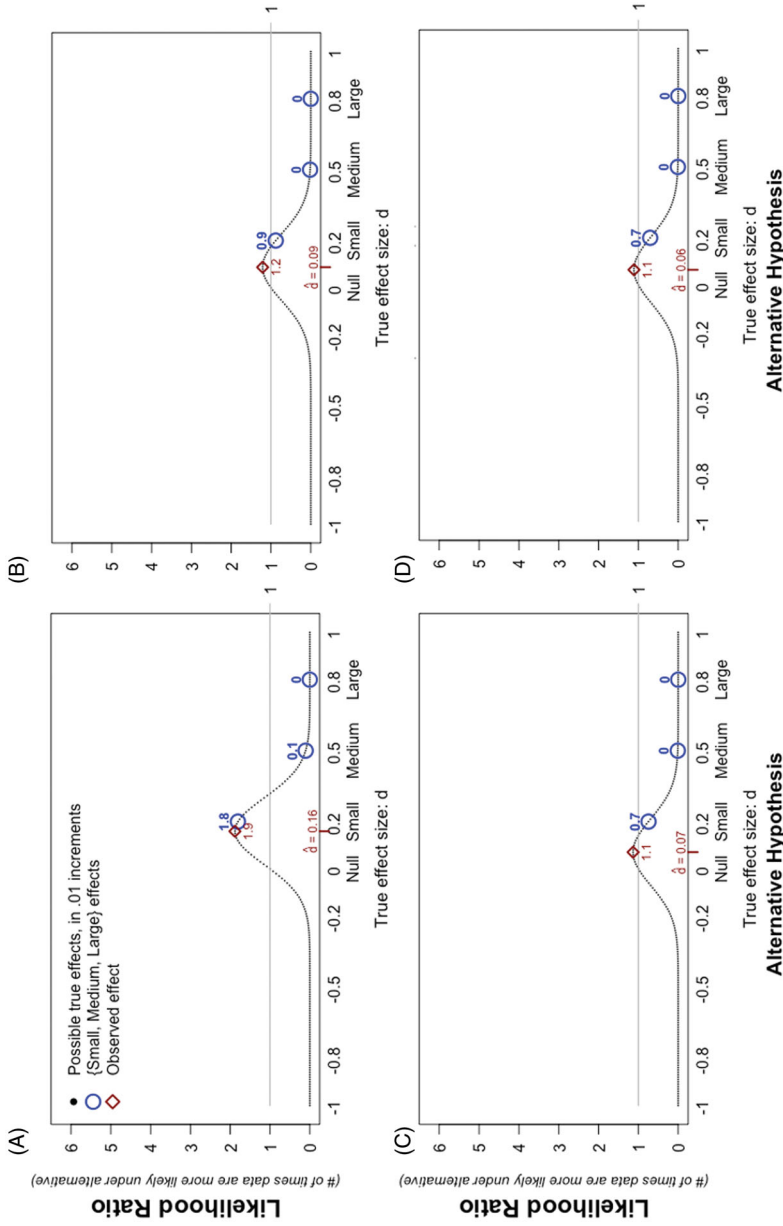


Figure 2. The likelihood of observing the data if the true effect size is small, medium, or large, compared to the null hypothesis of no effect of testing optimality on quantity of free (Panel A) and cued recall (Panel B), and accuracy of free recall and cued recall (Panels C and D respectively).

high confidence were better calibrated in their confidence judgments in non-optimal than in optimal sessions. A follow-up experiment further confirmed this tendency: overall confidence–accuracy relationship was stronger in non-optimal than in optimal sessions (Experiment 2; Yaremenko et al., 2021).

In another study by Nowack and Van Der Meer (2018), morning-type participants managed to efficiently counteract the negative effect of non-optimal testing on the performance in a semantic analogy task. The pupil dilation data suggested that participants achieved this by strategically allocating the limited attentional resources in non-optimal sessions. Interestingly, our exploratory analyses point in a similar direction: in non-optimal sessions, stronger morningness preference in our participants was associated with a larger number of details reported and higher accuracy in cued recall statements. No such association was found in optimal sessions, suggesting that morning types may have engaged in strategies aimed at compensating for cognitive decline at circadian troughs.

Another important aspect of the methodology of our experiment concerns the retrieval instructions. Consistent with recommendations for eliciting information from eyewitnesses, we emphasised the importance of being as complete and accurate as possible and discouraged guessing. Such instructions can induce a more conservative response strategy appropriate to the eyewitness memory contexts (Fisher & Geiselman, 1992; Koriat & Goldsmith, 1998). No equivalent of such instructions was present in the protocols of previous studies that showed synchrony effects in memory performance. Emphasising the importance of accurate memory reports may have further encouraged our participants to engage in cognitive strategies aimed at counteracting negative effects of non-optimal testing, reinforcing a more meticulous approach to providing memory reports than in previous synchrony effect studies. Our design does not fully allow us to confirm the role of retrieval instructions

in our findings. Future studies can test this possibility by experimentally manipulating the type of retrieval instructions used (conservative versus lenient) and measuring participants' confidence in the details they report.

One limitation of our study is that encoding and retrieval took place in the same experimental session. This design did not allow us to assess the effects of non-optimal testing on encoding and retrieval differentially. Future studies may address this issue by separating the two memory stages into different testing sessions and manipulating testing optimality for each of them separately, for example by employing a Testing Optimality (optimal versus non-optimal) \times Memory Stage (encoding versus retrieval) design. Measuring participants' self-reported sleepiness in the testing sessions can potentially explain some variability in the data; future research may therefore include measures such as the Stanford Sleepiness Scale (Hoddes et al., 1972). Additionally, the synchrony effect model we used does not allow us to disentangle the effects of circadian rhythm from sleep-related factors that could potentially mask circadian fluctuations in alertness and performance. Future research can employ highly controlled protocols to obtain a clearer picture of the interrelations of the circadian and sleep-related contributions to memory performance in eyewitnesses (e.g. K. Wright et al., 2006). However, these protocols test participants under highly artificial conditions, which might limit the applicability of the findings to the real situations to which eyewitnesses are exposed.

Additionally, our findings are limited to situations when other encoding and retrieval instructions are optimal. Real-life eyewitnesses may often encode events under less favourable conditions, such as insufficient lighting (Wagenaar & Van Der Schrier, 1996), suboptimal distance (Lindsay et al., 2008; Loftus & Harley, 2005) or short exposure duration (Memon et al., 2003). Some of these factors may be more pronounced at circadian arousal

troughs, pointing out numerous perspectives for future research. Finally, elderly eyewitnesses are generally poorer eyewitnesses (see Fitzgerald & Price, 2015, for a recent meta-analysis) and are known to have strong time-of-day preferences (Adan et al., 2012; Monk & Buysse, 2014; Monk et al., 1991). Future studies can test the possibility that obtaining testimony during circadian peaks may partially compensate for this age-related decline in performance in older eyewitnesses.

To conclude, our study found no support for the hypothesis that the alignment of time of day with chronotype can affect accuracy of eyewitness statements. It is possible that time-of-day optimality effects in recall performance in eyewitnesses might have a smaller effect size than we anticipated and, therefore, would require larger sample sizes to be detected than the sample in our experiment. It is important to test this possibility in future research for better understanding of metacognitive regulation of memory performance in eyewitnesses under suboptimal conditions. From the applied perspective, however, such a small effect size is unlikely to warrant the need for respective policy changes.

Ethical standards

Declaration of conflicts of interest

Sergii Yaremenko has declared no conflicts of interest

Melanie Sauerland has declared no conflicts of interest

Lorraine Hope has declared no conflicts of interest

Ethical approval

All procedures involving human participants were in accordance with the ethical standards of the Ethics Review Committee Psychology and Neuroscience of Maastricht University and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

Informed consent was obtained from all individual participants included in the study.

Note

1. We conducted the main analyses including only participants with originally suggested rMEQ scores and obtained analogous findings.

Acknowledgements

We would like to thank Marlot van Kempen and Anne Lausberg for their help in piloting the materials and collecting data. Open data and analytic code are available online at https://osf.io/5zvdv/?view_only=4fd8e48956564ca9a3b286263019fd72

Funding

This research is supported by a fellowship awarded from the Erasmus Mundus Joint Doctorate Program The House of Legal Psychology with Framework Partnership Agreement 2013-0036 under Specific Grant Agreement 2016-1339 to Sergii Yaremenko.

ORCID

Sergii Yaremenko  <http://orcid.org/0000-0003-3386-5812>

Melanie Sauerland  <http://orcid.org/0000-0002-7401-1365>

Lorraine Hope  <http://orcid.org/0000-0002-0012-8165>

References

- Adan, A., & Almirall, H. (1991). Horne & Östberg morningness-eveningness questionnaire: A reduced scale. *Personality and Individual Differences*, 12(3), 241–253. [https://doi.org/10.1016/0191-8869\(91\)90110-W](https://doi.org/10.1016/0191-8869(91)90110-W)
- Adan, A., Archer, S., Hidalgo, M., Di Milia, L., Natale, V., & Randler, C. (2012). Circadian typology: A comprehensive review. *Chronobiology International*, 29(9), 1153–1175. <https://doi.org/10.3109/07420528.2012.719971>

- Blagrove, M. (1996). Effects of length of sleep deprivation on interrogative suggestibility. *Journal of Experimental Psychology: Applied*, 2(1), 48–59. <https://doi.org/10.1037/1076-898X.2.1.48>
- Brackmann, N., Sauerland, M., & Otgaar, H. (2019). Developmental trends in lineup performance: Adolescents are more prone to innocent bystander misidentifications than children and adults. *Memory & Cognition*, 47(3), 428–440. <https://doi.org/10.3758/s13421-018-0877-6>
- Brown, A. L., & Smiley, S. S. (1977). Rating the importance of structural units of prose passages: A problem of metacognitive development. *Child Development*, 48(1), 1–8. <https://doi.org/10.2307/1128873>
- Caci, H., Deschaux, O., Adan, A., & Natale, V. (2009). Comparing three morningness scales: Age and gender effects, structure and cut-off criteria. *Sleep Medicine*, 10(2), 240–245. <https://doi.org/10.1016/j.sleep.2008.01.007>
- Correa, Á., Alguacil, S., Ciria, L. F., Jiménez, A., & Ruz, M. (2020). Circadian rhythms and decision-making: a review and new evidence from electroencephalography. *Chronobiology International*, 37(4), 520–541. <https://doi.org/10.1080/07420528.2020.1715421>
- Czeisler, C., & Gooley, J. (2007). Sleep and circadian rhythms in humans. *Cold Spring Harbor Symposia on Quantitative Biology*, 72, 579–597. <https://doi.org/10.1101/sqb.2007.72.064>
- Di Milia, L., Adan, A., Natale, V., & Randler, C. (2013). Reviewing the psychometric properties of contemporary circadian typology measures. *Chronobiology International*, 30(10), 1261–1271. <https://doi.org/10.3109/07420528.2013.817415>
- Diges, M., Rubio, M. E., & Rodriguez, C. (1992). Eyewitness memory and time of day. In F. Losel, D. Bender, & T. Bliesener (Eds.), *Psychology and law. International perspectives* (pp. 317–320). W. De Gruyter.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/brm.41.4.1149>
- Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/bf03193146>
- Fisher, R. P., & Geiselman, R. E. (1992). *Memory enhancing techniques for investigative interviewing: The cognitive interview*. Charles C. Thomas.
- Fisk, A., Tam, S., Brown, L., Vyazovskiy, V., Bannerman, D., & Peirson, S. (2018). Light and cognition: Roles for circadian rhythms, sleep, and arousal. *Frontiers in Neurology*, 9, 56. <https://doi.org/10.3389/fneur.2018.00056>
- Fitzgerald, R., & Price, H. (2015). Eyewitness identification across the life span: A meta-analysis of age differences. *Psychological Bulletin*, 141(6), 1228–1265. <https://doi.org/10.1037/bul0000013>
- Frenda, S. J., Patihis, L., Loftus, E. F., Lewis, H. C., & Fenn, K. M. (2014). Sleep deprivation and false memories. *Psychological Science*, 25(9), 1674–1681. <https://doi.org/10.1177/0956797614534694>
- Goldstein, D., Hahn, C. S., Hasher, L., Wiprzycka, U. J., & Zelazo, P. D. (2007). Time of day, intellectual performance, and behavioral problems in Morning versus Evening type adolescents: Is there a synchrony effect? *Personality and Individual Differences*, 42(3), 431–440. <https://doi.org/10.1016/j.paid.2006.07.008>
- Halberg, F., Cornélissen, G., Katinas, G., Syutkina, E. V., Sothorn, R. B., Zaslavskaya, R., Halberg, F., Watanabe, Y., Schwartzkopff, O., Otsuka, K., Tarquini, R., Frederico, P., & Siggelova, J. (2003). Transdisciplinary unifying implications of circadian findings in the 1950s. *Journal of Circadian Rhythms*, 1(1), 2. <https://doi.org/10.1186/1740-3391-1-2>
- Hastie, R., & Kumar, P. A. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology*, 37(1), 25–38. <https://doi.org/10.1037/0022-3514.37.1.25>
- Hoddes, E., Dement, W., & Zarcone, V. (1972). The development and use of the stanford sleepiness scale (SSS). *Psychophysiology*, 9, 150.
- Horne, J. A., & Östberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology*, 4(2), 97–100. <https://doi.org/10.1177/0748730405285278d>
- Jeffreys, H. (1961). *Theory of probability* (3rd ed). Oxford University Press.
- Koriat, A., & Goldsmith, M. (1998). The role of metacognitive processes in the regulation of

- memory performance. In G. Mazzoni & T. O. Nelson (Eds.), *Metacognition and cognitive neuropsychology: Monitoring and control processes* (pp. 97–118). Lawrence Erlbaum Associates Publishers.
- Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139087759>
- Lindsay, R. C., Semmler, C., Weber, N., Brewer, N., & Lindsay, M. R. (2008). How variations in distance affect eyewitness reports and identification accuracy. *Law and Human Behavior, 32*(6), 526–535. <https://doi.org/10.1007/s10979-008-9128-x>
- Loftus, G. R., & Harley, E. M. (2005). Why is it easier to identify someone close than far away? *Psychonomic Bulletin & Review, 12*(1), 43–65. <https://doi.org/10.3758/bf03196348>
- May, C. P., Hasher, L., & Foong, N. (2005). Implicit memory, age, and time of day: Paradoxical priming effects. *Psychological Science, 16*(2), 96–100. <https://doi.org/10.1111/j.0956-7976.2005.00788.x>
- May, C. P., Hasher, L., & Stoltzfus, E. R. (1993). Optimal time of day and the magnitude of age differences in memory. *Psychological Science, 4*(5), 326–330. <https://doi.org/10.1111/j.1467-9280.1993.tb00573.x>
- Memon, A., Hope, L., & Bull, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. *British Journal of Psychology (London, England : 1953), 94*(Pt 3), 339–354. <https://doi.org/10.1348/000712603767876262>
- Monk, T. H., & Buysse, D. J. (2014). Chronotype, bed timing and total sleep time in seniors. *Chronobiology International, 31*(5), 655–659. <https://doi.org/10.3109/07420528.2014.885981>
- Monk, T. H., Reynolds, C. F., 3rd, Buysse, D. J., Hoch, C. C., Jarrett, D. B., Jennings, J. R., & Kupfer, D. J. (1991). Circadian characteristics of healthy 80-year-olds and their relationship to objectively recorded sleep. *Journal of Gerontology, 46*(5), M171–175. <https://doi.org/10.1093/geronj/46.5.m171>
- Morgan, D., Tamminen, J., Seale-Carlisle, T., & Micks, L. (2019). The impact of sleep on eyewitness identifications. *Royal Society Open Science, 6*(12), 170501. <https://doi.org/10.1098/rsos.170501>
- Moritz, S., Woodward, T. S., & Rodriguez-Raecke, R. (2006). Patients with schizophrenia do not produce more false memories than controls but are more confident in them. *Psychological Medicine, 36*(5), 659–667. <https://doi.org/10.1017/S0033291706007252>
- Nowack, K., & Van Der Meer, E. (2018). The synchrony effect revisited: Chronotype, time of day and cognitive performance in a semantic analogy task performance in a semantic analogy task. *Chronobiology International, 35*(12), 1647–1662. <https://doi.org/10.1080/07420528.2018.1500477>
- Paine, S., Gander, P., & Travier, N. (2006). The epidemiology of morningness/eveningness: Influence of age, gender, ethnicity, and socio-economic factors in adults (30–49 years). *Journal of Biological Rhythms, 21*(1), 68–76. <https://doi.org/10.1177/0748730405283154>
- Petros, T., Beckwith, B., & Anderson, M. (1990). Individual differences in the effects of time of day and passage difficulty on prose memory in adults. *British Journal of Psychology, 81*(1), 63–72. <https://doi.org/10.1111/j.2044-8295.1990.tb02346.x>
- Puttaert, D., Adam, S., & Peigneux, P. (2019). Subjectively-defined optimal/non-optimal time of day modulates controlled but not automatic retrieval processes in verbal memory. *Journal of Sleep Research, 28*(4), e12798. <https://doi.org/10.1111/jsr.12798>
- Roediger, H., & McDermott, K. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803–814. <https://doi.org/10.1037//0278-7393.21.4.803>
- Roenneberg, T., Kuehnl, T., Juda, M., Kantermann, T., Allebrandt, K., Gordijn, M., & Merrow, M. (2007). Epidemiology of the human circadian clock. *Sleep Medicine Reviews, 11*(6), 429–438. <https://doi.org/10.1016/j.smrv.2007.07.005>
- Ryan, L., Hatfield, C., & Hofstetter, M. (2002). Caffeine reduces time-of-day effects on memory performance in older adults. *Psychological Science, 13*(1), 68–71. <https://doi.org/10.1111/1467-9280.00412>
- Sauerland, M., Krix, A., van Kan, N., Glunz, S., & Sak, A. (2014). Speaking is silver, writing is golden? The role of cognitive and social factors in written vs. spoken witness accounts. *Memory & Cognition, 42*(6), 978–992. <https://doi.org/10.3758/s13421-014-0401-6>
- Schmidt, C., Collette, F., Cajochen, C., & Peigneux, P. (2007). A time to think: Circadian rhythms in human cognition.

- Cognitive Neuropsychology*, 24(7), 755–789. <https://doi.org/10.1080/02643290701754158>
- Sim, J., & Wright, C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy*, 85(3), 257–268. <https://doi.org/10.1093/ptj/85.3.257>
- Stebly, N. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law and Human Behavior*, 21(3), 283–297. <https://doi.org/10.1023/A:1024890732059>
- Stebly, N. M. (1992). A meta-analytic review of the weapon focus effect. *Law and Human Behavior*, 16(4), 413–424. <https://doi.org/10.1007/BF02352267>
- van Kesteren, M. T. R., Ruitter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, 35(4), 211–219. <https://doi.org/10.1016/j.tins.2012.02.001>
- Wagenaar, W. A., & van der Schrier, J. H. (1996). Face recognition as a function of distance and illumination: A practical tool for use in the courtroom. *Psychology, Crime & Law*, 2(4), 321–332. <https://doi.org/10.1080/10683169608409787>
- Wright, A. M., & Holliday, R. E. (2007). Enhancing the recall of young, young-old and old-old adults with cognitive interviews. *Applied Cognitive Psychology*, 21(1), 19–43. <https://doi.org/10.1002/acp.1260>
- Wright, K., Hull, J., Hughes, R., Ronda, J., & Czeisler, C. (2006). Sleep and wakefulness out of phase with internal biological time impairs learning in humans. *Journal of Cognitive Neuroscience*, 18(4), 508–521. <https://doi.org/10.1162/jocn.2006.18.4.508>
- Yang, L., Hasher, L., & Wilson, D. (2007). Synchrony effects in automatic and controlled retrieval. *Psychonomic Bulletin & Review*, 14(1), 51–56. <https://doi.org/10.3758/bf03194027>
- Yaremenko, S., Sauerland, M., & Hope, L. (2021a). Eyewitness identification performance is not affected by time-of-day optimality. *Scientific Reports*, 11(1), 3462. <https://doi.org/10.1038/s41598-021-82628-z>
- Yaremenko, S., Sauerland, M., & Hope, L. (2021b). *Time-of-day optimality effects in the production of spontaneous false memories in morning and evening types*. Manuscript in preparation.

Appendix

Cued questions about the event

1. Describe any interactions the thief/thieves had with the other people in the film.
2. How did the thief/thieves get the opportunity to steal the wallet?
3. What was the victim doing when the thief/thieves stole the wallet?
4. How long did the theft last?
5. Were there any accomplices?
6. What did the thief/thieves do with the stolen wallet?
7. What did the victim do when she/he realized the wallet was stolen?
8. How many people did you see in the film?
9. Is there any other information you would like to share with us about the event that we have not asked you about?

Cued questions about the thief

1. How old was the thief? (Enter one number, not a range)
2. How tall was the thief in cm? (Enter one number, not a range)
3. Describe the thief's build.
4. Describe the thief's clothing.
5. Describe the thief's hair colour.
6. Describe the thief's hairstyle.
7. Describe the thief's face shape.
8. Did you notice any special features in the appearance of the thief?
9. Did the thief wear something on his/her head? If yes, what?
10. Did the thief wear glasses? If yes, what did they look like?