Time-of-Day Optimality Effects on Eyewitness Memory Performance

Sergii Yaremenko

This thesis is submitted in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy of The University of Portsmouth

September 2019

The House of
Legal Psychology

Supervisors:

Prof. dr. Lorraine Hope	*University of Portsmouth, United Kingdom*
Dr. Melanie Sauerland	*Maastricht University, the Netherlands*
Prof. dr. H.L.G.J. Merckelbach	*Maastricht University, the Netherlands*

Assessment Committee:

Dr. Paul Morris	*University of Portsmouth, United Kingdom*
Prof. dr. Graham Pike	*The Open University, United Kingdom*

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 and Specific Grant Agreement 2016-1339 to Sergii Yaremenko.

General Abstract

Our body clock, also known as the circadian rhythm, regulates alertness and activity levels across the day and determines periods of optimal cognitive performance. The exact timing of peak performance may vary due to inter-individual differences in the functioning of our body clock: Morning types perform best in the morning, whereas evening times reach their peak performance in the evening. This pattern is known as the synchrony effect. There are numerous benefits in scheduling cognitively demanding activities in congruence with individual performance patterns. Recall and recognition performance have been shown to be better at the optimal compared to the non-optimal hours of the day. However, the role of the synchrony effect in eyewitness memory performance received little attention in the literature.

In the research programme presented in the current thesis, we aimed to investigate the effects of time-of-day optimality on memory performance in eyewitnesses. Across three single- and multi-session experimental studies, we (i) investigated the effect of time-of day optimality on accuracy and informativeness of free narratives and answers to cued questions about the mock crime (Chapter 1); (ii) explored time-of-day optimality effects on lineup identification performance and its postdictors (Chapters 1 and 2); (iii) tested time-of-day optimality effects in face recognition performance and memory for sources in which faces were encountered (Chapter 3), and (iii) explored possible synchrony effect patterns in the formation of false memories (Chapter 1).

In **Chapter 2** (*N = 103),* we tested whether matching individual time-of-day preferences can be beneficial for accuracy and informativeness of eyewitness reports, accuracy of eyewitness identification decisions and postdictive value of confidence and decision times. We also investigated whether time-of-day optimality affects the formation of false memories. Time-of-day optimality did not affect performance in eyewitness reports and false memory rates. Identification accuracy in target-present lineups was unexpectedly higher at non-optimal compared to the optimal time of day. Highly confident choosers were significantly better calibrated in their confidence judgments at non-optimal compared to optimal hours. Decision times were predictive of accuracy only at the optimal time of day.

In **Chapter 3** *(N = 324),* we further investigated the possibility of the synchrony effect in identification accuracy and its posdictors. Results showed no significant differences in identification accuracy between optimal and non-optimal sessions. In line with findings reported in Chapter 3, confidence-accuracy relationship was stronger at the non-optimal time of day. Decision times were not predictive of accuracy.

In **Chapter 4** *(N = 91),* we tested the possibility of synchrony effects in face recognition performance and ability to discriminate between the contexts in which faces were encountered. Results showed no benefit in overall face recognition performance. Participants showed no benefit from optimal testing in terms of their ability to exclude familiar but irrelevant faces. These findings are novel in demonstrating that face recognition performance is not subject to synchrony effect patterns commonly reported in the literature.

Overall, the results of our experiments show no evidence for the synchrony effect in any of the investigated aspects of eyewitness memory performance. It can be concluded that in healthy young adult eyewitnesses, circadian troughs in cognitive performance are not sufficient to result in significant reduction of evidential value of testimony, providing that other encoding and retrieval conditions were optimal. Our data suggest that eyewitnesses can take the presence of factors that impair cognitive performance into account and adjust their confidence judgments appropriately.

Table of Contents

Declaration

Whilst registered as a candidate for the above degree, I have not been registered for any other research award. The results and conclusions embodied in this thesis are the work of the named candidate and have not been submitted for any other academic award.

Sergii Yaremenko

Word count: 37,610

List of Figures and Tables

Figures

Tables

Dissemination

Publications

Yaremenko, S., Sauerland, M., Hope, L. (2019) Circadian variations in eyewitness identifications: Identification performance is not affected by time-of-day optimality. *Manuscript submitted for publication, revisions invited from Memory.*

Yaremenko, S., Sauerland, M., Hope, L. (2019) Synchrony effects in accuracy of eyewitness memory reports. *Manuscript in preparation.*

Yaremenko, S., Sauerland, M., Hope, L. (2019) Time-of-day optimality and memory for faces: Circadian arousal, recognition and source memory performance. *Manuscript in preparation.*

Conference Presentations

Yaremenko, S., Sauerland, M., Hope, L. (June 2019) Chronotype and time-of-day effects on eyewitness identification performance. *Poster presentation at the thirteenth biennial conference of the Society for Applied Research in Memory and Cognition (SARMAC) in Cape Cod, USA.*

Yaremenko, S., Sauerland, M., Hope, L., Merckelbach, H. (June 2018) Chronotype and time-of day effects on eyewitness memory. *Presentation at the annual meeting for the European Association for Psychology and Law (EAPL), Turku, Finland.*

Yaremenko, S., Sauerland, M., Hope, L., Merckelbach, H. (May 2017) Chronotype and time-of day effects on eyewitness memory. *Presentation at the annual meeting for the European Association for Psychology and Law (EAPL), Mechelen, Belgium.*

Chapter 1: Introduction

Our body clock, also known as the circadian rhythm, is a biological mechanism that allows our body to adapt to cyclic changes occurring in the environment throughout the day (Halberg et al., 2003). This unique biological oscillator has an immense impact on all aspects of our functioning, from purely biological functions to various aspects of behaviour and cognition. It coordinates the timing of numerous systems in our organism with 24-hour changes in the external environment and keeps them in synchrony with each other (Hastings, Reddy, & Maywood, 2003).

Sleep and wake cycle may be among the most apparent manifestations of the work of our body clock. Some other fluctuations caused by the circadian rhythm are so inherent in our daily functioning that we barely pay any attention to them. They range from changes in appetite, body temperature and metabolism to daily fluctuations in our mood, arousal and ability to be at our best in physical and mental tasks we encounter throughout the day (Reppert & Weaver, 2002).

Research has demonstrated advantages to understanding how our internal body rhythms work and, whenever possible, making the timing of various activities congruent with this cycle. This "congruency" appears to affect numerous aspects of our functioning, from administering medications (e.g., Smolensky & Peppas, 2007) or planning sports trainings at optimum hours (Peek et al., 2017) to scheduling cognitively demanding activities at times of the day that facilitate top performance, such as later school time starts for adolescents (Kelley, Lockley, Foster, & Kelley, 2015).

The legal arena represents another domain where optimal cognitive performance may be critical. Eyewitness testimony often features as important evidence in criminal cases. Therefore, the performance of eyewitnesses can play a crucial role in the administration of the law, with eyewitness errors potentially leading to wrongful convictions and other miscarriages of justice (Clark, Benjamin, Wixted, Mickes, & Gronlund, 2015). Yet we know very little about how prior experimental research on circadian variations in arousal and cognitive performance translates into the eyewitness memory domain. For example, do eyewitnesses encode the details of crimes differently at different points of the day? Could the evidential value of identification decision vary as a function of the time of day when perpetrator's face was encoded? Can we help eyewitnesses provide accurate statements by conducting the interview at the peak of their cognitive performance? Are some hours of the day more optimal for administering lineups than others? These questions have received no attention in the research literature to date. The current thesis attempts to address this gap in the literature by taking the first steps in investigating circadian effects in eyewitness memory performance.

The Circadian Rhythm and Chronotype

Our heart produces around 108,000 beats per day with a rate that varies at different points of the day. The functioning and structure of cells involved in our stomach, liver and pancreas vary from hour to hour, allowing our digestive system to function properly across the day. Over 50 different hormones are produced and circulated in our body, ensuring the proper functioning of immune system and responses to stress, and making us feel sleepy at certain hours and alert at others. It is essential that these processes are coordinated with the external environment and each other. Luckily, a clock-like mechanism known as the circadian rhythm (from the Latin *circa*, meaning "around", and *diem*, meaning "day") leaves little room for chaos in this highly complex system (Glass, 2001). A group of neurons located in the suprachiasmatic nucleus of the hypothalamus acts as a central pacemaker, coordinating the functioning of multiple physiological, hormonal and behavioural systems (Refinetti, 2006; Reppert & Weaver, 2002). The circadian rhythm ensures that the timing of physiological and behavioural events is consonant with the demands the external environment poses at each specific moment of the day.

The circadian clock is entrainable, that is, it can be "set" by the external cues such as light and food intake cycles and social cues. However, in the absence of external time cues, for instance, in constant darkness, the circadian clock continues to maintain near-24-hour cycles, proving to be an endogenous, self-sustained biological clock that works with impressive precision (Borbély & Achermann, 1999; Gaggioni, Maquet, Schmidt, Dijk, & Vandewalle, 2014).

The exact timing of sleep, activity and performance phases can vary among individuals. These inter-individual differences in sleep and activity timings are known in the literature as morningness-eveningness preference (Horne & Östberg, 1976), or chronotype (Roenneberg, Wirz-Justice, & Merrow, 2003). The chronotype dimension is best described as a continuum (Natale & Cicogna, 2002). Evening types, who wake up and go to sleep late and prefer evening hours for their activity, are at one end of the continuum. Morning types, on the other hand, wake up and go to sleep earlier, and prefer to be active in morning hours (Horne & Östberg, 1976; Roenneberg et al., 2007). Inter-individual differences in the circadian rhythm are not limited to sleep habits: Chronotype also influences individual preferences in the timing of meals, determines hours of the day with peak hormone levels and body temperature, and affects a plethora of other factors of our daily functioning (for a review, see Adan et al., 2012).

Chronotype is most commonly studied with the help of self-report tools that measure 24-hour changes in subjective alertness and preferences in the timing of sleep, physical and cognitive activities. The Morningness-Eveningness Questionnaire (MEQ) constructed by Horne & Ostberg (1976) was the first such self-report tool and is still widely used in chronobiological research to measure of time-of-day preferences. The MEQ consists of 19 questions about participants' sleeping habits, alertness levels at different points of the day, and preferred times for engaging in physically and cognitively demanding tasks (e.g., "At what time in the evening do you feel tired and in need of sleep?"). The MEQ score has been shown to correlate with daily changes in melatonin and cortisol levels, body temperature (Baehr, Revelle, & Eastman, 2000; Horne & Ostberg, 1976), sleep habits and activity levels (Andrade, Benedito-Silva, & Menna-Barreto, 1992; Bailey & Heitkemper, 2001; Duffy, Rimmer, & Czeisler, 2001).

A reduced version of the questionnaire (rMEQ; Adan & Almirall, 1991), was developed as a shorter alternative to the original MEQ. It includes five items from the full MEQ, offering a convenient alternative to the somewhat lengthy full version of the questionnaire without significant loss inter-item correlation and validity (Natale, Esposito, Martoni, & Fabbri, 2006). The MEQ and its reduced version are the most commonly used tools to measure individual differences in the circadian rhythm (Adan et al., 2012).

Some of the other self-report tools used to measure chronotype are Munich Chronotype Questionnaire (Roenneberg et al., 2003), Diurnal Type Scale (Torsvall & Åkerstedt, 1980), and Composite Scale of Morningness (Smith, Reilly, & Midkiff, 1989). For a review of the psychometric differences in various questionnaires, see Di Milia, Adan, Natale, and Randler (2013).

The Circadian Rhythm, Sleep, and Memory Performance

The circadian rhythm can affect long-term memory performance via two distinct mechanisms. First, the circadian rhythm (together with homeostatic sleep pressure) regulates the timing of sleep (Van Dongen & Dinges, 2003), which is a pre-requisite to any effortful cognitive activity (Antonenko, Diekelmann, Olsen, Born, & Mölle, 2013; Cousins, Sasmita, & Chee, 2017). Next, sleep the night following encoding of new information plays an especially important role in memory functioning by promoting memory consolidation. Consolidation makes the newly encoded memory trace more stable and immune to interference. Problems with the sleep schedule that are caused by disruptions of our circadian clock (e.g., issues resulting from jetlag or shift work, the circadian rhythm disorder) may have detrimental effects on newly formed memories, making them unstable and more prone to interference (for a review, see Rasch & Born, 2013).

Another mechanism of circadian regulation of memory performance relates to the role of our body clock in maintaining cycles in alertness and arousal levels in our waking life. Generally, our cognitive performance tends to be better during periods of high alertness. We tend to be more alert during the day and sleepier at night; thus, by regulating the levels of arousal at different points of the day, the circadian rhythm determines hours that are optimal in terms of memory performance (Reppert & Weaver, 2002). The effect of such time-of-day optimality on eyewitness memory performance is the primary focus of the current research programme.

Not everyone reaches their peak at the same hours of the day, and cognitive performance varies is a function of alignment of the task timing with individual's circadian arousal peaks, which are determined by chronotype. This pattern is referred to as the synchrony effect: Individuals with morning preference are at their best in the morning hours, whereas evening types peak later during the day (May, 1999; May &

Hasher, 1998). The synchrony effect patterns have been demonstrated across a large variety of cognitive domains, such as attentional capacities and vigilance, inhibition of irrelevant responses, and performance on tasks that rely on short- and long-term memory (for a review, see Schmidt, Collette, Cajochen, & Peigneux, 2007).

Episodic memory in particular appears to show synchrony effect patterns across a wide range of memory tasks, including free recall, cued recall and recognition. To measure variations in free recall performance as a function of alignment of time of testing and individual's optimal performance timings, Petros, Beckwith, and Anderson (1990) recruited morning- and evening-type participants and tested them either at 9 AM, 2 PM or 8 PM. Participants studied prose passages and wrote down what they remembered in their own words. Morning-type participants produced significantly more idea units from the passages at 9 AM compared to the later sessions; the reported data indicate that the observed effect was of small size. Evening-type participants showed a similar tendency and reported more idea units in the evening session compared to the two earlier sessions, although the difference was not statistically significant.

In a similar manner, cued recall performance appears to be better at circadian peaks compared to circadian troughs. May, Hasher, and Foong (2005) compared performance on a stem completion task in the morning (8:00 AM and 9:00 AM) or in the early evening (between 5:00 PM and 6 PM) in younger morning-types and older evening-types (i.e. chronotype was fully confounded with age). First, they introduced the target words to participants in what appeared to be a pleasantness-rating task. Subsequently, their memory for these words was tested in a stem completion task. Performance was better at the optimal time of day in both chronotype groups:

Morning-type participants recalled more studied words in the morning, whereas evening types completed more stems correctly in the evening.

Recognition performance can also be susceptible to circadian variations in arousal. May, Hasher, and Stoltzfus (1993) tested morning-and evening-type participants either in the morning (8 AM or 9 AM) or in the late afternoon (4 PM or 5 PM). After encoding short stories, participants were presented with old and new sentences and had to indicate whether each of them had been present verbatim in one of the original stories. In evening-type participants, hit rates were higher and false alarms were lower at their optimal compared to the non-optimal time of day. In morning-type older adults, the hit rates at the optimal and the non-optimal time of day were similar, whereas false alarms increased in the non-optimal session. The corrected recognition scores, computed as hits minus false alarms, confirmed overall better performance at the optimal time of day in both chronotype groups, with a small to medium size of the observed effect.

Non-optimal testing can also boost unwanted cognitive biases, such as false memories in the famous Deese–Roediger–McDermott (DRM) paradigm. In this paradigm, participants encode thematically related items (e.g., nurse, hospital, ill, medicine, stethoscope). When their memory is tested, participants consistently report studying the associatively related critical lure (e.g., doctor), even though the lure has not been presented to participants (Cann, McRae, & Katz, 2011; Roediger & McDermott, 1995; Roediger, Watson, McDermott, & Gallo, 2001). Intons-Peterson, Rocchi, West, McLellan, and Hackney (1999) report a synchrony effect in false memory production in older participants, who showed higher memory rates at non-optimal compared to the optimal time of day.

These experiments are a part of a large body of research showing that the synchrony effect in memory performance appears to be a robust phenomenon observed in all basic memory tasks (for other examples, see Puttaert, Adam, & Peigneux, 2018; Ryan, Hatfield, & Hofstetter, 2002; Yang, Hasher & Wilson, 2007; Yoon, 1997). Circadian arousal is an important factor affecting the way we remember autobiographical events, with a small to medium average effect size, comparable, for instance, to the effect size of biased identification instructions (Steblay, 1997).

From a theoretical perspective, the synchrony effect is often construed as a cyclic variation in the amount of available cognitive resources (e.g., Nowack & Van Der Meer, 2018). We subjectively experience circadian arousal cycles as changes in energy levels throughout the day (Cariou, Galy, & Mélan, 2008; Thayer, 1987). The energy dimension of arousal is thought to reflect the amount of attentional resources available for any effortful mental activities we wish to engage in (Hirst & Kalmar, 1987; Necka, 1997).

Attention plays a crucial role in successful encoding of information, serving as a link between perception and memory (e.g., Chun & Turk-Browne, 2007). The amount of attention paid to an autobiographical episode determines the extent to which the event can be successfully recalled later (Craik, Govoni, Naveh-Benjamin, & Anderson, 1996)[1]. Encoding information in long-term memory is not merely the byproduct of paying attention to the stimulus: It largely depends on the degree of elaboration we engage in when the processing the incoming stimuli (Craik & Lockhart; 1972; Lockhart, 1992; Otten, Henson, & Rugg, 2001). Shallow processing leads to weak encoding, whereas deep, or more elaborate processing results in better

---

[1] This statement relates primarily to conscious, explicit recollection; the interrelation between and implicit memory is more complex (e.g., Mulligan, 1998).

memory for the studied material (Craik, 1999; Craik & Lockhart, 1972; Craik & Tulving, 1975). Elaborate processing occurs when we make connections between individual elements of the incoming information (Bousfield & Cohen, 1953; Tulving, 1962) or link the new information with pre-existing knowledge and experience (Craik; 2002; Hastie & Kumar, 1979). Optimal attentional capacities at circadian arousal peaks provide allow us to engage elaborate processing of incoming stimuli, which results in stronger encoding of an event (Valdez, 2019).

In a similar manner, retrieval from long-term memory is an effortful activity and can be sensitive to factors that negatively affect our cognitive resources (e.g., Fernandes & Moscovitch, 2000). To complete a difficult memory test, participants need to be able to pay proper attention to instructions and process them. Providing a detailed recollection of an incident can be a long and effortful task that consumes cognitive resources. Even mnemonic techniques that aid recollection require some amount of cognitive capacities. For instance, effort is necessary in order to mentally reinstate the context of a to-be-remembered event. Accurate retrieval may also require suppressing unwanted responses, an ability that largely relies on the available cognitive resources (e.g., Jacoby, Woloshyn, & Kelley, 1989; Kane & Engle, 2000).

Nonetheless, our understanding of mechanisms that underlie circadian variations in memory performance may not be complete. Recent neuroscientific studies demonstrated that cortical and hippocampal areas involved in memory functioning can manifest their own circadian oscillations that function as an autonomous peripheral clock complementing the central pacemaker located in superchiasmatic nucleus (Eckel-Mahan & Storm, 2009). It is likely that these intrinsic circadian rhythms in hippocampus play an important role in the time-of-day fluctuations in memory functioning. However, the underlying cellular and molecular mechanisms of these peripheral oscillations and their actual effect on memory performance are yet to be explored (Snider, Sullivan, & Obrietan, 2018).

The theoretical understanding of the mechanisms behind the synchrony effect are an interesting avenue for future research. While the primary focus of the current thesis lies in the applied field of eyewitness memory performance, exploring manifestations of circadian fluctuations in arousal across various contexts can contribute to overall understanding of the underlying mechanisms behind the synchrony effect.

Circadian Rhythms in Eyewitness Memory: Same Story, Different Characters?

Eyewitness testimony can be a two-edged sword in criminal justice system. An accurate and informative testimony may lead to a breakthrough in investigation (Graham, 2003), whereas memory errors can result in wrongful convictions (Clark et al., 2015). Therefore, the research into human memory and factors that affect it are of high importance to the legal arena.

Luckily, decades of sound scientific research into the functioning of human memory provided insights into how we remember autobiographical events. Traditionally, experimental protocols in memory research relied on variations of the so-called verbal-learning paradigm. In this model, participants study lists of words (or, alternatively, prose fragments, drawings, photographs of objects etc.). Memorizing the word list itself, as well as each specific item on the list, serves as an experimental model of encoding an autobiographical experience (Kihlstrom, 2009). Participants retain the studied material for various intervals. After that, experimenters test participant’s memory for the encoded experience. The generic memory tasks frequently used to test memory performance include free recall, in which participants are simply asked to report the experience they remember (“Write down the words that you studied earlier”), cued recall, in which participants are provided with extra cues that aid retrieval (“Use the words that you studied to complete these word stems”), or recognition (“Do you remember seeing this drawing?”). By manipulating conditions during encoding, retention or retrieval, researchers make inferences about the way various factors affect our memory.

The verbal-learning paradigm has proven to be an effective experimental model simulating human memory processes in laboratories of cognitive psychologies. Over decades of research that relied on variations of the verbal-learning paradigm, our understanding of fundamental principles of memory functioning has increased immensely (Kihlstrom, 1996). The synchrony effect literature is also based on analogous methods of studying memory: by comparing memory for words, sentences, prose passages and objects at the optimal and non-optimal hours of the day, we now know that memory is best during peaks of circadian arousal (Intons-Peterson et al., 1999; May et al., 1993; May et al., 2005; Petros et al.,1990; Puttaert et al., 2018; Ryan et al., 2002; Yang et al., 2007; Yoon, 1997).

Is it reasonable to assume that memory for crimes could be subject to similar circadian variations in performance? As a matter of fact, there are numerous examples of classic discoveries about memory functioning that map well onto the eyewitness memory field. Such factors as retention interval, exposure duration or the age of the witness are now widely acknowledged as important contributors to memory performance in eyewitnesses (see, e.g., Loftus, 1975; Wells & Olson, 2003).

However, not all laboratory findings can be applied to the eyewitness memory domain in such a straightforward manner. Conditions real eyewitnesses face can fundamentally differ from the protocol of a standard memory experiment. Witnessing a crime bears little resemblance to studying word lists or memorising prose passages.

Furthermore, the nature of the memory queries in the context of criminal proceedings significantly differs from memory tests in a memory lab, and the costs of memory errors in investigative contexts are incomparable to mistakes in recognizing the wrong word in a standard memory experiment.

To address these issues of ecological validity, a set of research methodologies known as the eyewitness memory paradigm was designed (Loftus, Miller, & Burns, 1978). Instead of memorizing word lists or text fragments, participants in an eyewitness memory studies may become mock witnesses to a (digitally recorded) staged crime. As a free recall test, they may be asked to provide free narratives of what they remember. Questions about the incident and people involved are an analogue of standard cued recall task (e.g., Brewer, Vagadia, Hope, & Gabbert, 2018). When asking participants to identify the perpetrator from a photographic lineup, experimenters essentially test their recognition memory (Loftus, 1979; Wells & Olson, 2003).

Eyewitness memory research has significantly advanced our understanding of how our general knowledge about memory functioning applies to the eyewitness memory domain. Factors that affect memory performance in eyewitnesses are traditionally classified into two broad categories: estimator variables, or factors outside of control of the criminal justice and law enforcement systems (e.g.. viewing conditions, age of the eyewitness, levels of intoxication), and system variables, which can be under control of the criminal justice system (e.g., the interviewing protocol, the lineup composition etc.; Wells, 1978; Wells, Memon, & Penrod, 2006).

As the circadian rhythm consistently appears to affect memory performance, time-of-day optimality could be among the potential estimator variables: memory for crime may vary as a function of the levels of circadian arousal in the eyewitness at the time of encoding. In a similar way, the circadian arousal could potentially act as a system variable: Perhaps it is beneficial to obtain eyewitness statements and administer lineup identifications at hours of the day when witnesses are at the peak of their cognitive performance. The methodology of previously conducted research allows us to draw only limited conclusions in these directions. The eyewitness memory paradigm, however, can offers adequate and valid methods of investigating the role of the circadian rhythm in eyewitness memory contexts.

To the best of our knowledge, only one study used the eyewitness memory paradigm to explore the possibility of circadian variations in eyewitness memory performance. In an experiment by Diges, Rubio, and Rodriguez (1992), mock eyewitnesses with either morning or evening chronotype were tested either at 10 AM or at 8 PM. They encoded a stimulus event depicting a traffic accident and provided free narratives and answers to cued questions about the witnessed event. Results were mixed, with an overall pattern of better performance in the morning compared to evening in both chronotypes. Diges et al. (1992) do not report mean MEQ scores of their, which would allow us to assess whether this tendency is related to the differences in the strength of time-of-day preference in each of the chronotype subgroups (e.g., it is possible that morning types showed a stronger preference for morning compared to the preference for the late hours in the evening types). These and other ambiguities and potential issues with the experimental method in this study, such as small sample size (with only 10 participants per experimental condition), ambiguous reporting of recall coding method, unclear exclusion criteria (e.g., caffeine consumption and sleep prior to testing) do not allow us to draw confident conclusions about these findings. Therefore, further research is needed to elucidate circadian variations in recall performance and other aspects of memory performance in eyewitnesses.

The Current Research

This programme of research presents the first attempts to test the synchrony effect patterns in eyewitness memory performance. Across three single- and multi-session experiments, we investigated whether (1) eyewitnesses provide more complete and accurate free narratives and answers to cued questions about the crime-related event at their optimal compared to the non-optimal time of day (Experiment 1); (2) eyewitnesses make more accurate lineup identification decisions at the optimal compared to the non-optimal time of day (Experiments 1 and 2); (3) face recognition performance is higher at circadian peaks compared to circadian troughs (Experiment 3); and (4) eyewitnesses are better at discriminating faces they encountered at the crime scene from faces that were encoded in unrelated contexts (Experiment 3). Given the onerous nature of the recruitment of morning- and evening-type participants and testing them within limited timeframe, we also pursued a secondary interest in examining whether the false memory rates are higher at non-optimal compared to the optimal time of day (Experiment 1). To the best of our knowledge, these are the first attempts to investigate these effects in the context of eyewitness memory performance.

Summary of the Thesis Chapters

Chapter 2: Circadian variations in eyewitness memory performance. This chapter presents an initial attempt to investigate time-of-day optimality effects across three domains relevant to eyewitness memory performance. For this purpose, we recruited one-hundred-and-three participants with morning and evening chronotype and tested each participant in the morning and in the evening. In each session, participants encoded a digitally recorded stimulus event that depicted a staged theft. At test, they provided free recall narratives about what they had seen, answered cued questions about the incident and the appearance of the people involved, and identified the individuals they saw in the stimulus event from target-present and target-absent lineups. We hypothesized that participants would provide more accurate free narratives and answers to cued questions when tested at their optimal compared to the non-optimal time of day. We also expected higher identification accuracy in the optimal compared to the non-optimal sessions. We hypothesized that non-optimal testing would weaken the postdictive value of confidence and decision times. Finally, we administered a visual version of the DRM paradigm, expecting that non-optimal testing would result in higher of false memory rates.

Accuracy in free narratives and answers to cued questions and false memory rates in the DRM task did not vary across the optimality conditions. Unexpectedly, participants showed higher identification accuracy in target-present but not target-absent lineups at non-optimal compared to the optimal time of day. Target selection rates were lower in optimal sessions, while foil selections did not vary as a function of testing optimality. The decision time-accuracy relationship commonly found in choosers was diminished at the non-optimal time of day. The confidence-accuracy relationship was not affected by testing optimality, with some evidence for superior calibration of confidence judgments at non-optimal compared to optimal sessions.

Chapter 3: Chronotype and time-of-day effects on eyewitness identification accuracy and its postdictors. Following our unexpected findings showing superior identification performance at non-optimal compared to the optimal time of day, we conducted Experiment 2 to further investigate synchrony effects in lineup identification performance. We tested three-hundred-and-twenty-four participants with morning or evening time-of-day preference recruited with the help of the Amazon MTurk platform. Participants were tested either at their optimal or the non-optimal time of day. After encoding a staged stimulus event, participants identified people involved in the depicted incident from target-present and target-absent lineups. We were interested in testing whether our surprising findings from Experiment 1 would replicate in a more diverse MTurk sample. Results showed no significant differences in identification accuracy between optimal and non-optimal sessions. Interestingly, evening-type participants were significantly more accurate in their identification decisions compared to morning types. Confidence was a stronger of accuracy in choosers at non-optimal compared to the optimal time of day. Decision times were not predictive of accuracy, likely due to the specifics of online testing environment.

Chapter 4: Circadian effects on face recognition and source memory performance. Long-term memory performance is generally better at the optimal compared to the non-optimal time of day. However, circadian fluctuations in face recognition performance received no attention in the literature to date. In Experiment 3, we used the face recognition paradigm to test possible synchrony effects in face recognition and source memory performance. Ninety-one morning- and evening-type participants were tested either in the morning or in the evening. Participants were presented with two sets of face stimuli in two contexts. One set of faces was embedded in a crime-related scenario, whereas the other set was presented in a neutral scenario. We hypothesised that general recognition performance would be better at the optimal compared to the non-optimal time of day. We also expected participants to be better at excluding familiar but irrelevant stimuli in the optimal compared to the non-optimal sessions. Results showed that overall recognition performance was unaffected by testing optimality. When asked to exclude familiar but irrelevant faces, participants showed only slight benefit from non-optimal testing that was not statistically significant. Although the current findings show no evidence for the effects of circadian arousal on face recognition and source memory for faces, future research is necessary to explore the underlying mechanisms of time-of-day effects in face recognition and source memory performance in vulnerable witnesses.

Chapter 5: General Discussion. In the final chapter, we attempt to synthesise the results obtained across all experiments within the research programme. The discrepancies between our findings and previous research are analysed from the viewpoint of divergence of methodological approaches to studying eyewitness memory performance from generic memory research methodology used in prior studies. An important line of discussion revolves around the problem of integrating findings obtained in generic laboratory experiments into applied contexts, such as eyewitness memory domain. The chapter outlines exciting directions for future research, such as the study of unorthodox synchrony effects in processing of faces and the effect of non-optimal testing on postdictors of eyewitness identification performance. We outline initial conclusions regarding the role circadian variations in arousal play in the eyewitness memory field, with an extra emphasis on the importance of replication of our findings due to the novelty of the research line. Finally, we discuss limitations of the current research programme and provide methodological suggestions for future study of the circadian effects in eyewitness memory performance.

Chapter 2: Circadian Variations in Recall and Identification Performance

Abstract

The circadian rhythm regulates arousal levels throughout the day and determines optimal periods for engaging in mental activities. Individuals differ in the time of day when they reach their peak: Morning-type individuals are at their best in the morning and evening types perform better in the evening. Performance in recall and recognition of verbal and pictorial non-facial stimuli is known to be superior at individual's circadian peak. Here, we tested the idea that eyewitness memory performance varies as a function of testing optimality.

One-hundred-and-three morning- and evening-type participants were tested both at their optimal and the non-optimal time of day. In each of the two testing sessions, participants viewed a stimulus film depicting a staged crime and provided free narratives of the stimulus event, gave answers to cued questions, and identified the individuals involved in the incident from target-present and target-absent lineups. We expected participants to provide more accurate statements and make more accurate identification decisions at the optimal compared to the non-optimal time of day. We also hypothesized that the confidence-accuracy and decision time-accuracy relationships would be undermined at the non-optimal time of day. Additionally, we administered a visual version of the Deese–Roediger–McDermott (DRM) paradigm. We expected participants to show higher false memory rates when tested at non-optimal compared to the optimal time of day.

Memory performance in free narratives and answers to cued questions did not vary as a function of time-of-day optimality. In terms of identification accuracy, participants were unexpectedly more accurate at non-optimal compared to the optimal time of day in target-present but not target-absent lineups. The decision time-accuracy relationship was undermined for choosers, when tested at the non-optimal time of day,

whereas the confidence-accuracy relationship was not affected by testing optimality. The production of spontaneous false memories in the DRM paradigm was not affected by time-of-day optimality. Further research is needed to elucidate the underlying mechanisms of circadian variations in eyewitnesses' decision-making processes.

Introduction

Life on earth revolves around a 24-hour daylight cycle, and most organisms on our planet, including humans, function in synchrony with this rhythm (Kyriacou & Hastings, 2010). This means that the timing of various processes and events in our body is not random; rather, they are synchronized with the cyclic changes occurring in external environment. The underlying clock-like mechanism responsible for these changes is known as the circadian rhythm (from the Latin *circa*, meaning "around", and *diem*, meaning "day") and is regulated by a group of neurons located in the suprachiasmatic nucleus of the hypothalamus. It coordinates the functioning of multiple physiological, hormonal and behavioural systems, allowing the organism to adapt to daily changes in the environment by ensuring the proper timing of a large number of physiological and behavioural events (Gaggioni, Maquet, Schmidt, Dijk, & Vandewalle, 2014). Some of the processes regulated by the circadian clock in humans include variations in body temperature, cardiac activity, hormone levels and metabolism (Borbély & Achermann, 1999).

The circadian clock also dictates the levels of physiological arousal at each specific moment of the day to maintain the optimum timing of rest and activity periods. To perform this important function, the circadian rhythm interacts with another process known as homeostatic sleep pressure (Van Dongen & Dinges, 2003). The sleep pressure phenomenon refers to the fact that arousal levels decrease and propensity to sleep increases as the day progresses. In lay terms, the more time we spend awake, the more tired and sleepy we feel. Circadian process and homeostatic process work together in regulating melatonin and cortisol levels to determine the propensity to physical and mental activities and sleep at each specific moment of the day (Borbély, Daan, Wirz-Justice, & Deboer, 2016).

Not everyone's body clock runs to the same timings; on the contrary, people widely differ in their preferred time of the day for sleep and activity, as determined by their chronotype, or time-of-day preference (Horne & Östberg, 1976; Levandovski, Sasso, & Hidalgo, 2013). Morning types, often referred to as 'larks', prefer waking up at early hours and find it difficult to stay awake in the evening. Evening types, or 'owls', prefer to go to sleep in the late hours and have difficulties getting up early in the morning. Intermediate types show no strong morning or evening preference (Adan et al., 2012).

Cognitive performance largely depends on whether or not the actual time of day is aligned with individual time-of-day preference (Schmidt, Collette, Cajochen, & Peigneux, 2007). Morning types reach their functional peak in the morning, as opposed to evening types, who are at their best in the evening hours. This phenomenon is known as the synchrony effect and it has been shown to affect inhibition of distractors, non-relevant thoughts and unwanted responses (May, 1999; May & Hasher, 1998), automatic application of stereotypes and other judgmental heuristics (Bodenhausen, 1990), implicit memory performance (May, Hasher, & Foong, 2005), and accessibility of information from semantic memory (Anderson, Petros, Beckwith, Mitchell, & Fritz, 1991).

This synchrony effect is also observed in long-term memory performance. Petros, Beckwith, and Anderson (1990) investigated whether matching or mismatching time of testing to participants' time-of-day preference affected immediate recall of prose passages, Morning types recalled significantly more idea units at 9 AM (optimal) compared to the afternoon and evening sessions (non-optimal), whereas no optimality effect was observed in evening-type participants. In another experiment, participants encoded a series of paragraph-length stories and then performed a verbatim sentence recognition task. The performance of evening participants improved from morning to afternoon, whereas performance of morning types was better in the morning, compared to the afternoon (May, Hasher, & Stoltzfus, 1993).

Overall, the body of research on circadian variations in cognitive performance suggests that time-of-day optimality is an important factor that can affect cognitively demanding activities. The legal arena represents a domain where optimal cognitive performance may be critical. Eyewitness testimony and lineup identifications often feature as important evidence in criminal cases. Therefore, the performance of eyewitnesses can play a crucial role in the administration of the law, with memory errors potentially leading to wrongful convictions and other forms of miscarriages of justice (Clark, Benjamin, Wixted, Mickes, & Gronlund, 2015). Yet the possible effects of time-of-day optimality on eyewitness memory performance have received no attention in the research literature to date.

Prior evidence in support of time-of day effects in long-term memory suggests that eyewitness memory performance may be subject to typical synchrony effect patterns. However, the nature of the stimuli and tasks employed in the previous studies showing the synchrony effect in memory performance is quite different from tasks and stimuli real-life eyewitnesses encounter. Specifically, participants in previous studies showing synchrony effect in memory performance had to encode word lists, sentences, short stories and pictures, as opposed to details of a crime-related event and people involved in it. As such, the question of how the prior findings translate into eyewitness memory field remains open.

The current experiment was designed as the first attempt towards investigating the possible time-of-day effects on eyewitness memory performance. We used a mock eyewitness identification paradigm, in which participants watched stimulus films depicting a staged crime and were subsequently tested using recall and recognition tasks relevant to the eyewitness context. First, we asked participants to provide free recall narratives about the event. Then all participants completed a set of cued questions about the event and the people involved. Based on past research showing superior recall during circadian peaks (Petros et al., 1990), we hypothesized that participants would provide more complete and accurate responses when tested at the optimal compared to the non-optimal time of day. Participants were also asked to identify the individuals they saw in the film from lineups. Based on prior evidence showing superior recognition performance during circadian peaks (Intons-Peterson, Rocchi, West, McLellan & Hackney, 1999; May et al., 1993), we expected participants to make more accurate identification decisions at the optimal compared to the non-optimal time of day.

Additionally, we investigated how chronotype synchrony affects the postdictive value of postdictors of identification accuracy, namely post-decision confidence and decision times. Typically, accurate choosers make their decisions with more confidence than inaccurate choosers, whereas such associations between confidence and accuracy do not exist for nonchoosers (Brewer & Wells, 2006; Dunning & Stern, 1994; Sauerland & Sporer, 2007; Sporer, Penrod, Read, & Cutler, 1995). There are reasons to expect that circadian asynchrony may have detrimental effect on the predictive value of confidence. For instance, confidence can be less predictive of accuracy in situations when encoding, retention and retrieval conditions are less optimal (Deffenbacher, 1980; Sauer & Hope, 2016). We tested the idea that chronotype asynchrony may be among such conditions. We predicted that testing participants at their preferred times would strengthen the confidence-accuracy relationship in choosers. More specifically, we expected confident choosers to be more accurate than non-confident choosers and this relationship to be stronger at the optimal compared to the non-optimal time of day. We did not expect to observe this relationship for non-choosers.

Similarly to post-decision confidence, research consistently shows a negative relationship between identification accuracy and response time in choosers but not non-choosers (Brewer & Wells, 2006; Dunning & Stern, 1994; Sauerland & Sporer, 2007). We were interested in whether testing optimality can act as a moderator of decision-time-accuracy relationship. To our knowledge, only one study investigated the effect of testing optimality on long-term memory access speed (Anderson et al., 1991). In this experiment, participants with morning or evening preference were asked to make judgements about word pairs. The testing sessions were scheduled at 9:00, 14:00 and 20:00. Results showed a typical synchrony effect pattern, that is, reaction times decreased across the day for evening types and increased for morning types. Based on these findings, we hypothesised that inaccurate choosers will take longer to make their decisions than accurate choosers, but even more so at non-optimal compared to the optimal time of day.

Time-of-day optimality has also been shown to affect the formation of spontaneous false memories induced with the Deese–Roediger–McDermott (DRM) paradigm (Roediger & McDermott, 1995). In this paradigm, participants encode thematically related items (e.g., bedroom, pillow, rest, dream) that are associatively related critical lure(s) (e.g., sleep). Though the critical lures are not actually presented to participants, it is a robust finding that participants claim to remember these thematically related items at rates comparable to true memory rates (Cann, McRae, & Katz, 2011; Roediger, Watson, McDermott, & Gallo, 2001). Intons-Peterson et al.

(1999) tested the effect of time-of-day optimality on the formation of false memories using both verbal and pictorial stimuli. Non-optimal testing resulted in higher false memory rates compared to optimal testing in older but not younger adults. One limitation of a study by Intons-Peterson et al. (1999) is that age was fully confounded with chronotype, that is, all morning-type participants were older adults and all evening types were younger adults. We were interested in testing whether a similar pattern of results would be obtained in a sample consisting of both morning- and evening-type young adults.

Method

Participants

We conducted power analysis for a repeated-measures within-factors ANOVA with G*Power v3.1 (Faul, Erdfelder, Buchner & Lang, 2009; Faul, Erdfelder, Lang & Buchner, 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, effect size *f* of 0.18, alpha error probability of .05 and a power of .95 were used, which resulted in a required sample size of 103. We continued the pre-screening until we achieved the planned sample size. To achieve the planned sample size, we pre-screened 203 individuals who expressed interest in participation in the experiment for their circadian typology using the short form of Morningness-Eveningness Questionnaire (rMEQ; Adan & Almirall, 1991). One-hundred-and-three pre-screened participants whose rMEQ score was 12 and lower (evening types) and 17 and above (morning types) were invited to participate in the main experiment (15 male, 87 female, 1 unspecified; age 18 to 58, *M* = 22.6, *Mdn* = 22 years). The sample consisted of university students (*n* = 98) and members of the general population (*n* = 4). About half of the sample consisted of evening-type

(54.3%, $n = 56$, $M_{\mathrm{rMEQ}} = 9.82$, $SD_{\mathrm{rMEQ}} = 1.88$) and morning-type participants (45.6%, $n = 47$, $M_{\mathrm{rMEQ}} = 18.6$, $SD_{\mathrm{rMEQ}} = 1.50$). Participants were native Dutch ($n = 56$), German ($n = 26$) or English ($n = 21$) speakers. We recruited only Caucasian participants to avoid cross-racial bias in the identification task (Meissner & Brigham, 2001; Wilson, Berstein, & Hugenberg, 2016).

Design

The experiment used a repeated measures design, with time-of-day optimality serving as predictor. Each participant was randomly assigned to be tested both at 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. In terms of recall performance, the number of correct details provided by participants and accuracy rates served as dependent variables.

Participants also made seven identification decisions across two testing sessions. Overall, 54.8 % of lineups presented to participants were target-present and 45.2% were target-absent. The distribution of target-presence across the optimality conditions was such that 52.8 % of lineups presented in the optimal testing condition and 56.7 % of lineups in the non-optimal condition were target-present. Identification accuracy (accurate versus inaccurate), post-decision confidence (on an 11-point scale ranging from 0-100%, with intervals marked in 10% steps), and decision times (measured in ms) served as outcome variables.

Finally, in each session, participants were presented with a DRM task designed to induce spontaneous false memories. In each session, participants were presented with two of the four DRM stimuli. The six possible combinations of pairs of stimuli were counterbalanced across the optimality conditions. The proportions of ‘yes’ responses to presented items (hit rates), false alarms to non-presented critical lures and non-presented non-related controls served as the outcome variables.

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) developed Horne and Östberg (1976). Both MEQ and its reduced version are commonly used to assess individual differences in diurnal preferences (Adan et al., 2012; Di Milia, Adan, Natale, & Randler, 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 four 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 originally suggested of ≤ 11 for evening types and ≥ 18 for morning types Adan and Almirall (1991). In our decision to adopt more lenient cutoffs, we were guided by the view that chronotype is a continuous rather than a dichotomous construct, as well as the debate around the arbitrariness of the cutoffs suggested by the authors of questionnaires measuring diurnal preferences (Caci, Deschaux, Adan, & Natale, 2009).

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, $r(100) = .92, p < .005$.

Stimulus films. Two different stimulus films depicting the theft of a wallet were used. The films differ in the details of the event, the environment, and the actors. Film 1 (adapted from Sauerland, Krix, van Kan, Glunz, & Sak, 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, Sauerland, & Otgaar, 2018) the theft occurs in a university communal area. Three amateur actors (2 male, 1 female, 21-26 years old) appear in Film 2.

Photo lineups. Six-person target-absent and target-present simultaneous lineups were constructed for each of seven targets (Film 1; one thief, one victim, two bystanders; Film 2: one thief, one victim, one bystander). Each lineup included six shoulder-up photographs that were arranged in two rows of three pictures and labelled 1 to 6. The target positions for the perpetrator, victim, bystander 1, and bystander 2 lineups in Film 1 were 4, 5, 4, and 3, respectively. The target positions for the perpetrator, victim, and bystander in Film 2 were 3, 6, and 4, respectively. In target-absent lineups, the targets were replaced with photographs of another foil. The effective lineup sizes (Tredoux's *E*s) were established in a pilot study, in which 19 to 38 mock witnesses (total $N = 219$) were presented with a description of the target and chose a lineup member that best matches the description (Tredoux, 1999). Tredoux's *E*s ranged from 3.6 to 5.6. The proportions of identification of the targets in the pilot study ranged from 4.7% to 30% (not exceeding 7.8% in the perpetrator lineups), with critical ratio for difference from chance ranging between -3.1 and 1.3 (not exceeding -2.01 in the perpetrator lineups).

DRM stimuli. To induce spontaneous false memories, we used four black-and-white pictures depicting scenes of a classroom, a beach, a funeral and a wiretap operation adopted from Moritz, Woodward, and Rodriguez-Raecke (2006). The pictures were designed in such a way that each of the scenes contained 12 items that are typical for the context of the respective scene. For instance, a scene of a classroom depicted a teacher, students, a blackboard etc. Further, eight thematically related items were excluded from each scene, e.g., a pointer or a sponge, were excluded from a scene of a classroom. These eight items served as critical lures in the recognition task. These materials have been successfully used to induce spontaneous false memories (e.g., Peters et al., 2012; Otgaar, Howe, Peters, Smeets & Moritz, 2014).

The scenes differed in emotional valence (neutral, positive, negative and delusional paranoid). However, we did not have reasons to predict that time-of-day optimality would affect the formation of spontaneous false memories differently depending on the emotional valence of the scene, nor was there a significant interaction with testing optimality when this factor was entered as a predictor into a statistical model. Therefore, this factor will not be discussed any further.

Procedure

Pre-screening. Participants were recruited using advertisements on a university notice board and by actively handing out flyers. We told participants that the study concerned the effects of eating and caffeine-consumption habits on long-term memory in individuals with morning and evening time-of-day preference. Prospective participants who believed they had a morning or an evening preference contacted the research team, and were asked to fill out the pre-screening questionnaire to determine their eligibility to participate. The questionnaire consisted of five rMEQ items intermixed with filler questions concerning participants' eating habits included to provide additional support for the cover story. Participants with rMEQ scores ≤ 12 (evening types) or ≥ 17 (morning types) were invited to participate in the main experiment. They were instructed to exclude alcohol or caffeine-containing products and sleep a minimum of six hours prior to testing.

Main experiment. Participants were tested individually on two separate occasions, which were scheduled in the morning (between 8:00 and 10:00) and in the evening (between 19:00 and 21:00). The two testing sessions took place with an interval of a minimum of 36 hours to avoid possible fatigue following a non-optimal session. The protocol for the two sessions was analogous, except where specifically indicated.

First, participants watched one of the two stimulus films. We instructed participants to pay close attention to every detail, as they would later be asked to act as an eyewitness. Immediately after the film, 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. Participants were given 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 in the incident. First, they were presented with nine cued recall questions about the event (e.g., "Describe any interactions the thief/thieves has 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 informed them that they would be asked 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 etc. Blocks of questions about each of the persons involved in the event were presented separately in the following order: thief, victim, bystander 1, and (in case of 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).

All the recall instructions and questions were presented in participants' native language and participants were given a choice to provide responses in their native language. For the complete list of cued questions, see Appendix A.

Free reports and answers to cued questions were followed by a 30-minute interval to avoid potential verbal overshadowing effect in identification tasks (Meissner & Brigham, 2001). During this interval, participants filled in either the Pittsburgh Sleep Quality Index (Session 1; Buysse, Reynolds, Monk, Berman, & Kupfer, 1989) or demographic questionnaires (Session 2), followed by a visual version of the DRM paradigm task (both sessions).

In the DRM task, we informed participants that they would see two pictures and asked them to attend closely to each of them as they would be asked questions about them later. Each of the pictures was then presented for 40 secs. The order of the pictures was randomized. After encoding the pictures, participants engaged in a filler task (tetris) for 5 minutes. Next, participants were presented with two lists of items, presented verbally, corresponding to each of the presented pictures and were asked whether they saw each of the items in the respective picture. Each of the lists contained 24 items, 12 of which were present in the original picture, eight were non-presented thematically related lures[2], and four were non-presented unrelated controls. The items were presented in a random order.

Finally, participants were asked to identify each of the persons they saw in the film from a six-person simultaneous lineup. The perpetrator lineup was always presented first. The remaining two or three lineups, respectively, appeared in a random order. Participants were informed that the targets may or may not be present in the lineups and were encouraged to select the "Not present" option if they were not sure or didn't know. Decision times for each identification decision were recorded automatically. After each decision, participants indicated their confidence.

At the end of the second session, participants filled out the full version of the MEQ as an extra control of the chronotype classification. Participants received either gift vouchers worth € 27.50 or participation credit in return their participation and were debriefed via email upon the completion of data collection.

Recall coding

Following Wright and Holliday (2007), we developed a scoring template for both stimulus events. Each information unit from the events was categorized 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)

---

[2] Retrospectively, we realized that for one of the pictures, participants were presented with seven instead of eight critical lures due to a program error. To control for this, the proportion of correct responses rather than their number was entered as the outcome variable in the analyses.

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 in free narratives 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 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. Inter-rater reliability coefficients (kappas) ranged from $\kappa = .70$ to $\kappa = .87$, *p*s < .001, indicating substantial to almost perfect strength of agreement (Sim & Wright, 2005). When computing total accuracy rates, each detail provided by participants was counted once across free and cued recall.

Results

Data analyses. We conducted one-way repeated measures ANOVAs to ascertain the effect of time-of-day optimality on quantity of details provided by participants in recall tasks and respective accuracy rates, as well as on hit and false alarm rates in DRM tasks. Lineup identification outcomes were analysed with the help of Generalized estimating equations (GEE) models. GEE models are able to handle correlated data structures and thus can be employed to analyse binary outcomes collected from repeated measurements of the same individual (Zeger, Liang, & Albert, 1988). The GEE model allowed us to include all seven measurements from each participant in the same model. An alpha level of .05 was used for all inferential analyses.

One participant showed an inconsistent classification according to rMEQ (the score of 12, i.e., an evening type) and the full version of the questionnaire, where he scored 57 (almost reaching the morningness cut-off of 59). Data for this participant were excluded from analyses. Two participants failed to return to laboratory for the second session. We excluded their data from the analyses of performance in recall and DRM tasks but included their lineup identification responses from the session they attended in the GEE analyses.

Effect of time-of-day optimality on quantity of details and accuracy rates in recall tasks. There was no significant effect of time-of-day optimality on a number of correct details reported in free recall narratives, F (1, 99) = 0.14, p = .714, $\eta_p^2 = .001$, $M_{Opt} = 72.04$, $M_{Non\text{-}opt} = 70.28$, or answers to cued questions, F (1, 99) = 1.45, p = .232, $\eta_p^2 = .01$, $M_{Opt} = 56.45$, $M_{Non\text{-}opt} = 55.60$. To compute accuracy rates in free recall narratives and answers to cued questions, we divided the total number of accurate items reported by the total number of items reported (correct items + incorrect items + fabricated items, computes for each task separately). Testing optimality did not have a significant effect on accuracy rates in free narratives, F (1, 99) = 0.84, p = .363, $\eta_p^2 =.008$, or answers to cued questions, F (1, 99) = 0.25, p = .618, $\eta_p^{2=}$ .003. In fact, accuracy rates in optimal and non-optimal sessions were strikingly similar (FR $M_{Opt} = .95$, $M_{Non\text{-}opt} = .95$; CR $M_{Opt} = .87$, $M_{Non\text{-}opt} = .86$). Mean correct details, incorrect details and accuracy rates at the optimal and the non-optimal time of day across the recall tasks can be found in Table 1.

Table 1

Mean number of correct and incorrect details reported by participants and overall quantity and accuracy rates in free recall and cued recall tasks at the optimal and the non-optimal time of day, standard deviations in parentheses

Details		Optimal		Non-optimal	
		M	*(SD)*	*M*	*(SD)*
Free recall	Correct	72.04	18.83	70.28	19.04
	Incorrect	3.08	2.88	2.79	2.48
	Quantity	75.98	19.91	73.75	19.86
	Accuracy rate	.95	.04	.95	.04
Cued recall	Correct	56.45	11.03	55.60	10.65
	Incorrect	8.35	3.28	8.54	3.81
	Quantity	64.98	11.54	64.37	11.64
	Accuracy rate	.87	.05	.86	.05

Effect of time-of-day optimality on identification accuracy, choosing and other lineup outcomes. We conducted GEE analyses to establish the effect of time-of-day optimality (optimal versus non-optimal) and target presence (target-present versus target-absent) on identification accuracy (accurate versus inaccurate). We deleted non-significant terms stepwise. The final model included a significant Optimality x Target-Presence interaction, Wald $\chi^2(1) = 3.93$, $p = .048$. Simple slopes analyses revealed that participants made more accurate decisions at non-optimal compared to the optimal time of day in target-present, Wald $\chi^2(1) = 5.92$, Exp(B) = 1.60, $p = .015$, but not target-absent lineups, Wald $\chi^2(1) = 0.17$, Exp(B) = 0.92, $p = .680$. That means, in target-present lineups, participants were 1.6 times more likely to make an accurate decision at non-optimal compared to the optimal time of day (corresponding to a small effect size, Chen, Cohen, & Chen, 2010). This finding was contrary to our expectations.

Following up on our unexpected findings, the GEE analyses were run in order to test whether lineup choosing rates, that is, rates of positive identifications as opposed to lineup rejections, were affected by testing optimality. Optimality (optimal versus non-optimal) and target-presence (target-present versus target-absent) were entered as predictors in the model; choosing (selection versus rejection) served as the outcome. The analysis revealed a significant Optimality x Target Presence interaction, Wald $\chi^2(1) = 4.78$, $p = .029$. Simple slopes analyses showed that participants were less likely to make a positive identification decision at the optimal compared to the non-optimal time of day in target-present, Wald $\chi^2(1) = 14.21$, Exp(B) = 0.52 , $p < .001$, but not target-absent lineups, Wald $\chi^2(1) = 0.06$, Exp(B) = 1.05, $p = .802$[3]. The odds ratio of 0.52 corresponds to a small effect size (Chen et al., 2010).

Table 2 displays frequencies and proportions of different lineup outcomes for each of the lineups. The tendency to perform better at the non-optimal time of day was present (to a varying extent) across all target-present lineups except for victim lineup in Film 1, where participants performed overall poorly (29% correct at the optimal and 21% correct at the non-optimal time of day). In terms of perpetrator identifications, which are of primary applied interest, participants performed better when identifying the thief from Film 1 (77% correct at the optimal and 81% correct at the non-optimal time of day) compared to the thief in Film 2 (35% and 45% correct respectively). This difference may be linked to the differences in the effective size of the lineups (Tredoux's *E*'s were 4.3 and 5.6 respectively).

Finally, we tested whether optimality affected choosing rates for targets as opposed to foils differentially. For this purpose, we conducted two GEE analyses with

---

[3] Analyses of identification performance in solely perpetrator lineups showed no effect of time-of-day optimality. Therefore, I report the GEE models as they include all measurements collected.

target and foil selections (selection versus no selection) as the outcome variables. There was a significant effect of optimality on target selection: Participants were more likely to select the target at non-optimal compared to the optimal time of day, Wald $\chi^2(1) = 5.75$, Exp(B) = 1.58, $p = .016$, while the effect of optimality on foil selections was not statistically significant, $p = .488$.

Table 2

Frequencies and proportions of different lineup outcomes at the optimal and the non-optimal time of day

Lineup	**Optimal**		**Non-optimal**	
	Frequency	Proportion	Frequency	Proportion
Thief				
Target-present				
Targets choices	19	21.3%	11	15.3%
Foil choices	20	22.5%	15	20.8%
False rejections	50	56.2%	46	63.9%
Total	89	100%	72	100%
Target-absent				
Correct rejections	58	67.4%	44	64.7%
Foil choices	28	32.6%	24	35.3%
Total	86	100%	68	100%
Victim				
Target-present				
Target choices	55	59.2%	41	61.2%
Foil choices	19	20.4%	14	20.9%
False rejections	19	20.4%	12	17.9%
Total	93	100%	67	100%
Target-absent				
Correct rejections	51	62.2%	32	43.8%
Foil choices	31	37.8%	41	56.2%
Total	82	100%	73	100%
Bystander				
Target-present				
Target choices	45	54.8%	41	56%
Foil choices	19	23.2%	16	22%
False rejections	18	22%	16	22%
Total	82	100%	73	100%
Target-absent				
Correct rejections	39	41.9%	27	40.3%

Foil choices	54	58.1%	40	59.7%
Total	93	100%	67	100%

Sensitivity and response bias. In light of the unusual way testing optimality affected choosing rates in target-present lineups, we drew upon the signal detection theory approach (Green & Sweets, 1966) to further clarify the observed effects. Signal detection theory is widely used in the recognition literature to understand the mechanisms behind distinguishing novel items from those that were encountered before. For this purpose, signal detection analyses isolate hit and false alarm rates to compute two independent factors that affect recognition performance. *Discrimination accuracy* (d') is a measure that can be used to distinguish signals (in identification task signal refers to responses to targets) from noise (responses to foils). A value of zero indicates zero ability to distinguish targets from non-targets. *Response bias* (c), on the other hand, refers to the threshold for deciding that participants have seen the target before. In the context of eyewitness identification, a negative c value indicates a bias towards making a selection from a lineup, whilst a positive c value shows a bias towards rejecting the lineup.

We tested the two d' and c values for optimal versus non-optimal conditions collapsed across all lineups. Discriminability was slightly, though non-significantly higher in non-optimal, $d'_{\text{Non-Opt}} = 1.72$, compared to optimal condition, $d'_{\text{Opt}} = 1.34$, $G = 0.11$, $p = .915$. Response bias was $c_{\text{Non-Opt}} = 1.09$ in the non-optimal sessions, compared to $c_{\text{Opt}} = 0.94$ in the optimal session. We are unaware of a significance test for the c values, but on a descriptive level, the values do not to show a substantial difference between response bias levels in the two experimental conditions.

Figure 1 displays the ROC curves for optimal versus non-optimal testing conditions, with the diagonal line representing chance performance. The x axis represents false alarm rates computed as the proportion of foil identifications in target-absent lineups divided by six (i.e., the lineup size; Brewer & Wells, 2006), while the *y* axis represents the proportions of target selections in target-present lineups. Due to few observations in levels 0 to 40%, we collapsed these categories. The comparison of the curves confirms results of the GEE analyses: The curves shows higher hit rates in non-optimal compared to optimal sessions at all the cut-off levels, whereas the false alarm rates were similar for the two conditions.

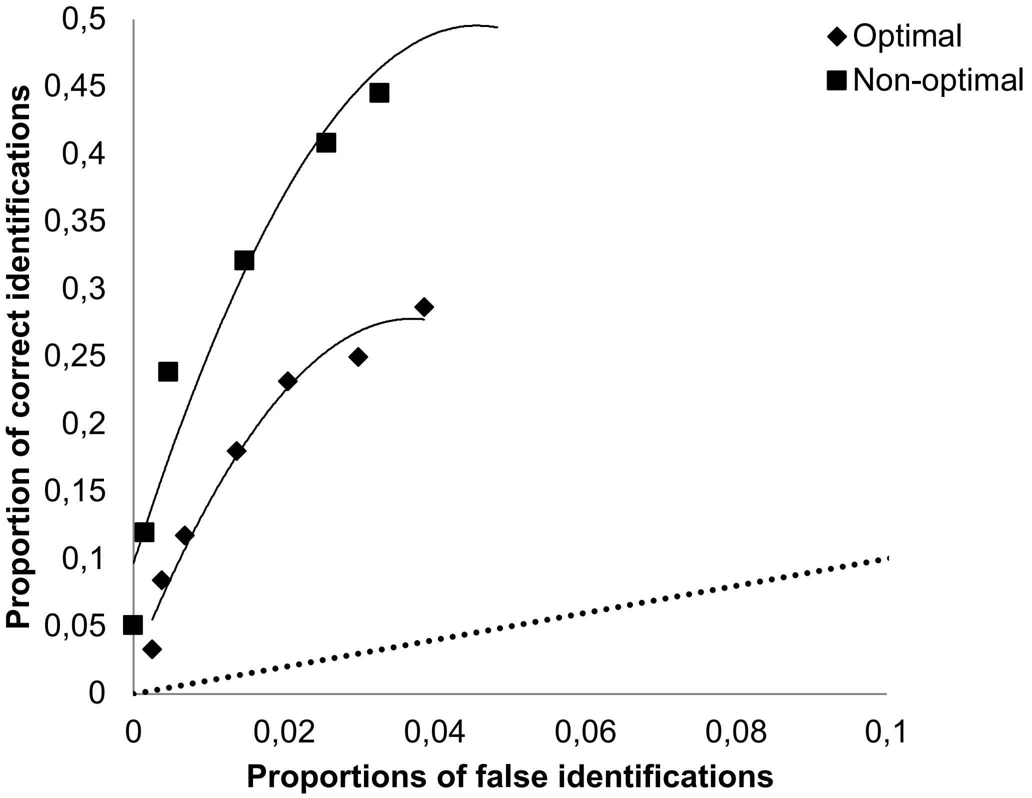


Figure 1. Receiver operating characteristics plots for the optimal and the non-optimal testing sessions across all lineups.

Effect of time-of-day optimality on the confidence-accuracy relationship. To establish the effect of time-of-day optimality on the confidence-accuracy relationship, we entered identification accuracy (accurate versus inaccurate), choosing (selection versus rejection), and time-of-day optimality (optimal versus non-optimal), as well as their two- and three-way interactions as predictors in the GEE model. Non-significant terms were deleted stepwise. The final model contained only the significant Identification Accuracy x Choosing interaction, Wald $\chi^2(1) = 13.80$, $p < .001$. Simple slopes analysis showed that accurate choosers were more confident than inaccurate choosers, Exp(B) = 3.09, Wald $\chi^2(1) = 50.40$, $p < .001$, whereas this was not the case for nonchoosers, Exp(B) = .82, Wald $\chi^2(1) = 1.09$, $p = .297$.

These findings replicate the common finding that accurate selections are made with more confidence than inaccurate selections (Sporer et al., 1995; Weber & Brewer, 2003, 2004). However, contrary to our expectations, this interaction effect was not moderated by testing optimality.

Calibration of confidence measures. Calibration analyses can be an informative way to explore on the association between the objective accuracy probabilities and subjective post-decision confidence measurements. There are several ways to assess how well eyewitnesses are calibrated in their confidence judgements. First, the proportion of accurate decisions for each confidence level can be plotted against the mean confidence for the respective level to create a calibration curve. Visual inspection of the curve allows to assess how well-calibrated participants were in each of the levels of confidence. Second, the calibration statistic (C) provides a quantitative reflection of the level of deviation from perfect calibration. It ranges from 0 (perfect calibration) to 1 (poorest calibration). Third, the over/underconfidence (O/U) statistic is a further indicator of how well-calibrated participants are. It varies from – 1 to + 1, with negative scores reflecting underconfidence and positive scores showing overconfidence. Finally, the normalized resolution index (*NRI*) allows to evaluate how good participant’s confidence judgments are at discriminating accurate from inaccurate decisions. It ranges from 0 (lowest resolution possible) to 1 (perfect discrimination; Weber & Brewer, 2004).

As expected and confirming the GEE analyses, the confidence-accuracy correlation was significant for choosers, $r(380) = .36$, $p < .001$, but not for nonchoosers, $r(319) = .04$, $p = .478$. Note the (nearly) large effect size for choosers that corresponds well with the effect reported in the meta-analysis (Sporer et al., 1995). Following Flowe et al. (2017), we collapsed the confidence categories into three categories (i.e., 0-40%; 50-70%; 80-100%) to provide more stable estimates for each confidence category. The proportion of accurate decisions for each of the collapsed categories was plotted against the weighted mean confidence for that category to create the confidence-accuracy calibration curve.

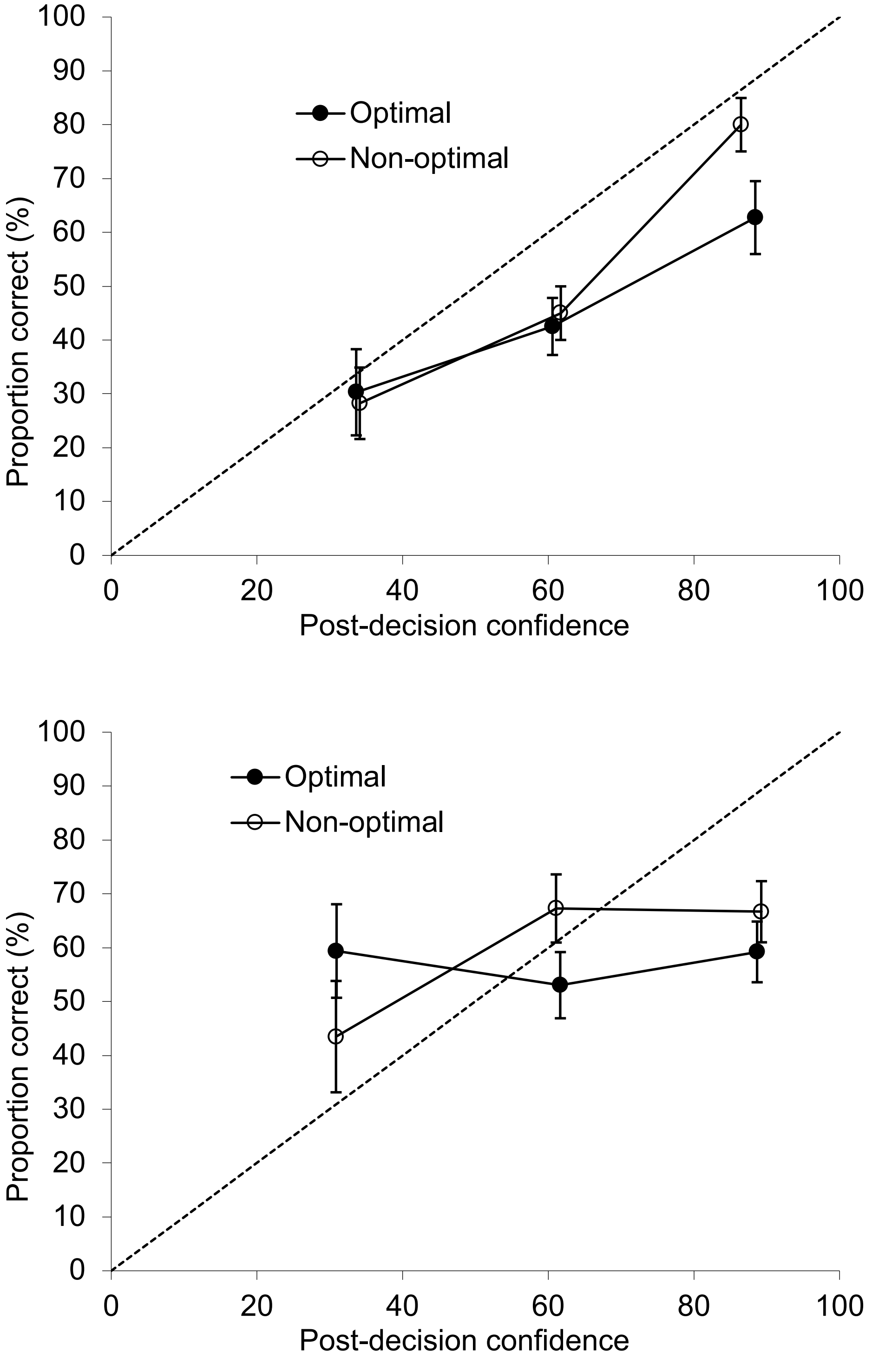


Figure 2. Decision confidence at the optimal and the non-optimal time of day for choosers (top panel) and non-choosers (lower panel).

Figure 2 (top panel) displays the confidence-accuracy calibration curves for choosers at the optimal (number of observations *n* = 171) and non-optimal (*n* = 211) time of day. The comparison of confidence intervals reveals that calibration in the last category (80–100%) was significantly better at non-optimal compared to the optimal time of day. Figure 2 (lower panel) shows the confidence-accuracy calibration curves for nonchoosers at the optimal (*N* = 174) and non-optimal (*N* = 147) time of day. Similarly to choosers, the comparison of confidence intervals shows that calibration in one of the confidence categories, namely the second category (40–70%), significantly differed at non-optimal as compared to the optimal time of day. Confidence intervals for the other confidence categories in choosers and nonchoosers overlap. Calibration statistics presented in Table 3 also show no significant differences between the optimality conditions.

Table 3
Calibration measures for optimal and non-optimal testing sessions split by choosers and nonchoosers

	Optimal		Non-Optimal	
	Measure	95% CI	Measure	95% CI
Choosers				
C	.036	.009; .064	.015	-.001; .032
O/U	.175	.104; .246	.112	.050; .174
NRI	.056	-.015; .126	.156	.061; .250
Nonchoosers				
C	.056	.020; .092	.028	.002; .054
O/U	.109	.027; .191	.063	-.019; .146
NRI	.037	-.034; .108	.004	-.015; .023

Confirming GEE analysis, these findings show no support for our hypothesis that postdictive value of confidence would be stronger at the optimal time of day. In fact, calibration analyses show some evidence in the opposite direction: Participants were better calibrated in some of the confidence categories in non-optimal compared to optimal sessions.

Effect of time-of-day optimality on the decision time-accuracy relationship. The decision time distributions showed significant positive skewness and kurtosis. Therefore, further analyses were conducted on log-transformed data (log base 10). To establish the effect of time-of-day optimality on the decision-time-accuracy relationship, we entered identification accuracy (accurate versus inaccurate), choosing (selection versus rejection), and time-of-day optimality (optimal versus non-optimal) and their interactions in the initial GEE model. Log-transformed decision times served as dependent variable. The model revealed a significant three-way interaction, Wald $\chi^2(1) = 4.03$, $p = .045$. Simple analyses of the Accuracy x Choosing interaction revealed a significant interaction effect at the optimal, Wald $\chi^2(1) = 6.19$, $p = .013$, but not at non-optimal time-of-day, Wald $\chi^2(1) < 0.01$, $p = .954$. Further analyses were performed on choosers and nonchoosers that were tested at the optimal time-of-day. The simple simple effect of accuracy was not significant for choosers, $b = 0.07$, Wald $\chi^2(1) = 3.31$, $p = .069$, or nonchoosers, $b = -0.07$, Wald $\chi^2(1) = 3.60$, $p = .058$. These results confirm our hypothesis, showing that non-optimal testing eliminated the postdictive value of decision times. However, the simple simple effect analyses fail to show a common finding that accurate chooser decisions are faster than inaccurate chooser decisions (Brewer & Wells, 2006; Dunning & Stern, 1994; Sporer et.al., 1995; Sauerland & Sporer, 2007, 2009).

Effect of time-of-day optimality on memory performance in the DRM paradigm. We conducted one-way repeated measures ANOVAs to determine whether testing optimality led to significant differences in proportions of yes responses to presented items, critical lures and non-presented non-related items. Recognition of presented items did not significantly differ between the optimal (M = .77, SD = .11) and non-optimal (M = .75, SD = .11) conditions, F (1, 99) = 1.50, p = .223, η_p^2 = .02. Similarly, proportions of Yes responses to critical lures were not affected by testing optimality, F (1, 99) = 0.05, p = .822, η_p^2 = .01, with M_{Opt} = .32, SD_{Opt} = .16 and $M_{Non\text{-}opt}$ = .32, $SD_{Non\text{-}opt}$ = .17. Finally, response rates to non-presented unrelated items did not significantly differ between optimal (M = .08, SD = .08) and non-optimal (M = .08, SD = .08) time of day, F (1, 99) = 0.03, p = .8, $\eta_p^2 < .01$. Net accuracy (calculated as the ratio of hit rates to hit rates plus false alarms to unrelated foils) did not significantly differ between the optimality conditions, F (1, 99) < 0.01, p = .979, $\eta_p^2 < .01$.

Due to potential contributions of "yea-saying" bias to hit and false alarm rates, it is common to transform hit and false alarm rates using the two-high-threshold correction, in which the proportion of yes responses to non-presented unrelated items is subtracted from hit rates and false alarms to critical lures (Otgaar et al., 2014; Snodgrass & Corwin, 1988). After we applied the correction, the pattern of findings was similar to the one obtained with untransformed scores. That is, testing optimality affected neither recognition of presented items, F (1, 99) = 0.91, p = .342, η_p^2 = .01(M_{Opt} = .69, SD_{Opt} = .14 and $M_{Non\text{-}opt}$ = .67, $SD_{Non\text{-}opt}$ = .14), nor responses to critical lures, F (1, 99) = 0.11, p = .746, η_p^2 = .02 (M_{Opt} = .24, SD_{Opt} = .15 and $M_{Non\text{-}opt}$ = .24, $SD_{Non\text{-}opt}$ = .17). We also obtained similar results when examining whether optimality affected performance for each scene separately (ps > .05). Mean hit rates, false alarms to critical lures and unrelated foils, and net accuracy across the four pictures can be found in Table 4.

Table 4

Proportions of Yes responses to presented items (hit rates), critical lures (false recognition) and non–presented non-related controls at the optimal and the non-optimal time of day

	Non-optimal					Optimal				
	Scene					Scene				
	Positive	Neutral	Negative	Delusional paranoid	Total	Positive	Neutral	Negative	Delusional paranoid	Total
Hit rate	.72	.69	.82	.77	.72	.73	.71	.84	.81	.73
Hit rate*	.47	.67	.80	.73	.47	.49	.68	.82	.77	.49
False recognition	.39	.33	.33	.26	.39	.37	.32	.32	.29	.37
False recognition *	.14	.313	.30	.23	.14	.13	.29	.30	.26	.13
Non-related	.25	.02	.03	.04	.25	.24	.02	.02	.03	.24
Net Accuracy	.67	.70	.74	.76	.67	.68	.71	.75	.75	.68

Note. * Rates corrected for response bias using two-high-threshold correction

Discussion

The current experiment investigated the effect of chronotype synchrony on eyewitness identification performance. To the best of our knowledge, this is the first attempt to investigate this effect in this context. Based on the previous findings that show quite consistent circadian variations in cognitive performance (e.g., Adan et al., 2012), we expected optimal testing to facilitate memory performance across an array of memory measures relevant to eyewitness domain, including recall, lineup identification and production of spontaneous false memories. Overall, our results show that previously reported findings on superior memory performance during circadian peaks do not translate into the eyewitness memory domain in the straightforward manner we anticipated.

In terms of recall performance, we expected participants to recall a higher number and proportion of accurate details at the optimal compared to the non-optimal time of day. Contrary to our expectations, accuracy rates did not significantly differ across the experimental conditions. In fact, accuracy rates in cued recall tasks were high at both optimal and the non-optimal time of day. In this respect, it is worthwhile noting that apart from the experimental manipulation, all other encoding and retrieval conditions were optimal. Participants were asked to pay close attention to the stimuli and were warned that they would have to act as eyewitnesses. Recall performance was measured immediately after participants encoded the stimulus event. We used recall instructions that discouraged guessing. In this experiment, we decided in favour of examining the mere effect of circadian arousal in the absence of other impairing factors to exclude the possibility of unpredicted interactions with the main manipulation. Future research can investigate whether time-of-day optimality effects in recall of eyewitnesses emerge with the use of more difficult cued questions or in the presence of additional suboptimal optimal factors, for example, an increased retention interval.

With regard to identification accuracy, we expected participants to make more accurate identification decisions at the optimal compared to the non-optimal time of day. Contrary to our expectations, the results demonstrated the opposite effect: Participants were more likely to make an accurate identification decision at their non-optimal compared to the optimal time of day. The effect was specific to target-present lineups; performance in target-absent lineups did not vary as a function of time of day. These findings run counter to the typical circadian patterns described in prior research: Typically, performance is better at the optimal compared to the non-optimal time of day (e.g., Intons-Peterson et al., 1999; May et al., 1993; May & Hasher, 1998; Ryan, Hatfield, & Hofstetter, 2002).

In order to obtain a more precise picture of how testing optimality affected our participants' choosing strategies in target-present lineups, we conducted additional analyses to examine whether our participants were less accurate at the optimal time of day because they selected fillers more often (i.e., due to the increase in false alarms), or because they incorrectly rejected the lineups even though the target was there (i.e., due to reduction in hit rates). These analyses showed that foil identifications were not affected by testing optimality. Instead, the drop in accuracy at the optimal time of day was the result of increased rejections of target-present lineups. Note that the size of the observed effect was small. We are also not aware of any other research that showed a decrease in hit rates during the circadian peak.

Several explanations may account for the discrepancy of our findings across both recall and recognition with the previous literature. First, our experiment differs from the previous studies in the nature of stimuli used. Participants encoded a stimulus film and were presented with a person identification task, as opposed to word, sentence, and picture recognition tasks used previously. Face recognition is traditionally believed to rely on cognitive processes that are relatively independent from other types of recognition memory (for a review, see Robotham & Starrfelt, 2017). It is possible that this dissociation also results in different circadian patterns in face recognition performance as opposed to verbal and pictorial recognition.

Another possible explanation may relate to the use of the repeated-measures design. We used a cover story to distract participant's attention from the main interest of our experiment. However, it was not possible to "blind" participants to the hour of the day when the testing occurred. Due to the fact that testing sessions took place twice at unusual hours of the day, participants may have become aware of the optimality manipulation, which may have potentially resulted in strategies aimed at compensating for the anticipated cognitive impairments in the non-optimal session, such as more careful stimulus encoding.

Overall, in terms of identification accuracy, our findings are unexpected in the light of theoretical framework and the observed effect size was small. It also remains unclear why the change in target selections was not accompanied by an effect in false alarms, as it is commonly observed in recognition memory performance studies (Wixted & Mickes, 2014). Therefore, the need for replication of this finding is evident.

We were also interested in whether testing optimality affected participants' post-decision confidence. Our hypothesis was that optimal testing will strengthen the predictive value of confidence often found in choosers but not non-choosers. We replicated the common finding that accurate selections are made with more confidence than inaccurate selections (Sporer et al., 1995; Weber & Brewer, 2003, 2004). However, contrary to our expectations, this interaction effect was not moderated by testing optimality. Calibration analyses confirmed that calibration statistic, over-under-confidence and resolution did not significantly differ between optimality conditions. In the meantime, visual inspection of the calibration curves showed that in some of the confidence categories, calibration was significantly better at the non-optimal compared to the optimal time of day; specifically, for highly confident choosers.

These findings are at odds with the optimality hypothesis proposed by Deffenbacher (1980), which predicts a higher postdictive value of confidence under optimal information-processing conditions. An alternative to Deffenbacher's optimality hypothesis was suggested by Palmer, Brewer, Weber, and Nagesh (2013). The authors point out that non-optimal conditions may not necessarily undermine confidence-accuracy calibration in situations when participants are aware of the presence of a factor that negatively affects their memory. In the context of our experiment, this could mean that due to their explicit morning or evening preference, our participants may have been aware of the possible cognitive impairments during the circadian troughs and taken this information into account in their metamemory judgements. This would explain the fairly good calibration in both optimality conditions with some evidence for superior calibration at non-optimal rather than the optimal time of day.

We also investigated whether testing optimality affected decision-time-accuracy relationship. We expected non-optimal testing to attenuate the predictive value of response latencies in lineup identifications in such a way that the decision-time-accuracy relationship in choosers would be weaker at non-optimal compared to the optimal time of day. Our results revealed that accurate choosers made faster decisions than inaccurate choosers, replicating a common result pattern (Brewer & Wells, 2006; Dunning & Stern, 1994; Penrod et.al., 1995; Sauerland & Sporer, 2007, 2009). The analyses also supported our prediction that the optimality would act as a moderator of this effect: Non-optimal testing completely eliminated the postdictive value of decision times. These findings suggest that the previously shown decrease of the speed of accessing information from long-term memory at the non-optimal time of day (Anderson et al., 1991) makes the difference in decision times between accurate and inaccurate choosers undetectable. If replicated under laboratory conditions, the findings of Experiment 1 would be interesting in terms of the predictive value of decision times in eyewitness identification decisions. They may also be of interest to memory researchers who rely on decision times as their outcome measure: It may be important to take into account that time of the day when the test is administered may confound the decision-time-based outcomes.

Our secondary interest lay in the effect of time-of-day optimality on the production of spontaneous false memories induced by the DRM paradigm. Results showed that testing optimality did not significantly affect hit rates or false alarms to critical lures. These findings are in line with results by Intons-Peterson et al. (1999), who reported higher false memory rates during circadian troughs in older but not younger adults. However, in the experiment by Intons-Peterson et al. (1999), age and chronotype were fully confounded (i.e., all younger participants were evening types). Therefore, our experiment extends previous knowledge on circadian variations in the production of spontaneous false memories by replicating this effect in sample consisting of both morning- and evening-type young adults. Combined with the data reported by Intons-Peterson et al. (1999), our results suggest that non-optimal testing does not produce significant differences in the production of spontaneous false memories in healthy young adults. Future research can focus on possible time-of-day variations in false memory rates under suboptimal conditions, such as divided attention or prolonged retention interval.

To conclude, creating optimal retrieval conditions for eyewitnesses is important for reducing identification errors and the associated potential of miscarriages of justice. This fact necessitates a closer examination of factors that contribute to memory performance in eyewitnesses. Our experiment contributes an initial understanding of the role of one of such factors, namely testing optimality, in eyewitness identification performance. Our results suggest that the effect of testing optimality on decision-making processes of eyewitnesses manifests itself in a more complex way than previously shown with non-eyewitness-identification paradigms: Non-optimal testing did not affect recall performance or production of spontaneous false memories but led to increased accuracy in target-present lineups. Further research is necessary to replicate these initial findings and further elucidate the underlying mechanisms of circadian variations in memory performance in the eyewitness.

Chapter 3: Disentangling Chronotype and Time-of-day Effects on Eyewitness Identification Accuracy

Abstract

The circadian rhythm regulates the timing of a large number of physiological and behavioural aspects of our functioning and determines optimal timings for engaging in different kinds of cognitive activities. Individuals differ in the timing of circadian peaks and troughs, as determined by their chronotype: Morning types show peak in their performance during the morning hours, whereas evening types are at their best in the evening. Prior research has shown that attentional capacities, inhibition of strong responses, and long-term memory are superior at the optimal compared to the non-optimal time of day.

We investigated the possibility that the circadian rhythm affects eyewitness identification outcomes. A total of 324 morning- and evening-type participants were recruited using the Amazon MTurk platform. Participants were randomly assigned to test sessions in the morning (7:30 AM – 9:00 AM) or in the evening (8:30 PM – 10:00 PM). Therefore, each participant was tested either at their optimal or the non-optimal time of day. Participants watched a stimulus film depicting a staged crime and were subsequently asked to identify each of the three individuals involved in the incident from six-person target-present and target-absent lineups. Based on previous studies showing an increase in recognition memory performance at time of the day that matches individual's chronotype, we expected participants to make more accurate identification decisions when tested at their optimal compared to the non-optimal time of day. We also expected optimal testing to strengthen the postdictive value of confidence and decision times.

Results showed that identification accuracy depended on chronotype alone, irrespective of whether the time of testing was optimal or non-optimal. Across all three lineups, evening types were 1.52 times more likely to make accurate

identifications compared to morning types. Contrary to our expectations, confidence-accuracy relationship was stronger at non-optimal compared to the optimal time of day. We did not observe a decision-time-accuracy relationship, possibly due to reduced control over the experimental setting in online environment. Findings from the current study shed new light on the way chronotype affects eyewitness identification accuracy.

Introduction

Daily variations in our physiology, behaviour and cognition are inherent in our daily lives. These changes do not happen chaotically. Instead, a central 'pacemaker' of our body, known as the circadian clock (Halberg et al., 2003) ensures that the numerous systems in our body function in synchrony with each other and with the external environment. Located in the suprachiasmatic nucleus of hypothalamus, the circadian clock maintains 24-hour cycles in heart rate and body temperature, hormone levels and metabolism, sleep propensity and alertness, to mention but a few (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, as some hours of the day are more optimal for engaging in cognitively demanding activities than others (Krishnan & Lyons, 2015; Kyriacou & Hastings, 2010). In a phenomenon is known as the synchrony effect, performance is better whenever the timing of the task is congruent with the circadian phase, and the other way around, the hours of the day that are incongruent with our internal body clock are associated with lowest performance (Schmidt et al., 2007).

Importantly, the circadian clock in some individuals is "set" in a slightly different way than in others. The so-called morning types, also known as 'larks', who 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 than most people (Horne & Östberg, 1976). This time-of-day preference is also referred to in the literature as the circadian typology, or chronotype (Roenneberg, Wirz-Justice, & Merrow, 2003). A relatively large proportion of the adult population (about 40%) belong to either a morning- or an evening-chronotype group (Adan et al., 2012).

Inter-individual differences in the functioning of the body clock can be an important factor determining what time of day that is most optimal in terms of cognitive performance. Across a wide range of cognitive domains, such as attentional capacities, working and long-term memory, inhibition of irrelevant responses, avoidance of stereotype-based responses and many others, morning types perform better in the morning, and, the other way around, evening types are better in the evening (Goldstein, Hahn, Hasher, Wiprzycka & Zelazo, 2007; May 1999; Nowack & Van Der Meer, 2018). In the same way, episodic memory is subject to the synchrony effect: Experimental research to date suggests that recall and recognition performance is better at circadian peaks as opposed to circadian troughs (May, Hasher, & Stoltzfus, 1993; Ryan, Hatfield & Hofstetter, 2002; Petros, Beckwith, & Anderson, 1990; Yoon, 1997).

In the meantime, time of day as a factor affecting memory performance of eyewitnesses has received little attention in the literature. If time of day determines information processing efficiency, could eyewitnesses be better at identifying the culprit from the crime they witnessed at their optimal as opposed to the non-optimal time of day? Or perhaps the hour of the day when the lineup is administered could affect identification outcomes? In order to explore possible time-of-day fluctuations in eyewitness identification performance, we conducted Experiment 1. The results were surprising: In target-present lineups, participants were more accurate at non-optimal compared to optimal sessions. The size of this effect was small and its direction contrasted previous literature. Additionally, the increase of hit rates was not associated with an increase in false alarms, as it is usually observed in recognition literature (Wixted & Mickes, 2014). Taken together, these issues highlighted the need for replication.

Thus, the current experiment was conducted to further investigate the effect of time-of-day optimality on eyewitness identification performance. We made several changes to the experimental procedure used in Experiment 1. We recruited participants via Amazon MTurk, which allowed us to collect a more demographically diverse sample. The increased statistical power resulting from a larger sample (cf. Experiment 1) made it possible to explore the possibility that time-of-day optimality manifests itself differently in individuals with morning as opposed to evening chronotype.

Next, we gave preference to a between-subjects design in the current experiment as opposed to within-subjects optimality manipulation in Experiment 1. We expected that this design would allow us to mask the main hypothesis of the study more successfully, fully eliminate the possibility of learning effects, and exclude random variations in performance associated with the use of two stimulus events. Additionally, we used a more thorough scheme of counterbalancing target positions and foil choices in the lineups to eliminate the possibility of obtaining a result that is attributable to some specifics of the retrieval stimuli (see Materials section below for detail).

Finally, in Experiment 2, we aimed to take a further step by introducing a memory bias in some of the experimental conditions. We were interested in testing the hypothesis that, if present in the lineup administration procedure, factors that are known to bias eyewitness's memory would be magnified by non-optimal testing. As means of introducing bias, we attempted to use the so-called mug shot exposure effect. In a nutshell, prior research showed that exposing eyewitnesses to mug shots before performing lineup identification may bias eyewitnesses' decisions. Under such conditions, eyewitnesses may base their decision on familiarity gained due to the mug shot presentation rather than on their memory from the actual event, thus increasing the likelihood of innocent suspect misidentifications (for a review, see Deffenbacher, Bornstein & Penrod, 2006). We hypothesised that such erroneous identifications due to mug shot exposure would be more likely to occur during circadian troughs compared to circadian peaks.

Experiment 2

Method

The experiment was pre-registered on the Open Science Framework. The registration form can be accessed at https://osf.io/kafe7/?view_only=bc5dfe1248f749678adb0cd8388c8b36.

Sample size. Power analysis for a two-tailed binomial logistic regression with G*Power v3.1 returned a required sample size of 310 (Faul, Erdfelder, Buchner & Lang, 2009; Faul, Erdfelder, Lang & Buchner, 2007). To achieve equal distribution of participants across the four experimental conditions, we tested two extra participants, which resulted in the planned sample size of 312.We used the following parameters: OR = 2.058; Pr(Y=1|X=1); H0 = 0.383; R^2 other X = 0.2, X distribution = Binomial, X parm $\pi = 0.5$. We based the odds ratio on the corrected recognition scores reported in a study by May et al. (1993). An alpha error probability of .05 and a power of .80 were used.

Participant pool. Participants were recruited with the help of Amazon Mechanical Turk (MTurk) platform. MTurk is an online crowd sourcing system that was developed with an aim to connect potential workers ("MTurkers") with the so-called requesters, who offer jobs or tasks that can be completed online. Researchers have been increasingly using the MTurk platform for the purposes of data collection due to the multiple advantages it offers for academic research (Mason & Suri, 2011; Paolacci & Chandler, 2014). MTurkers do not appear to be significant outliers in terms general demographics, and in some aspects the platform offers access to samples that are more representative compared to student samples (Buhrmester, Kwang & Gosling, 2011; Paolacci, Chandler, & Ipeirotis, 2010). The psychometric properties of responses collected on MTurk have been validated, and multiple laboratories managed to replicate some of the classical findings using the platform (e.g., Crump, McDonnell & Gureckis, 2013; Rand, 2012). Amazon MTurk offers opportunities for efficient and less costly data collection with reliable results. Combined with the fact that MTurkers are known to complete tasks around the clock, this has guided us in our decision to collect the data for Experiment 2 using the platform.

Data validity checks and inclusion criteria. Online testing can increase the rates of careless or partially random responses (Meade & Craig, 2012), an issue which may be of extra concern in studies that rely on Amazon MTurk platform in participant recruitment (e.g., Fleischer, Mead, & Huang, 2015). Therefore, we took a series of additional measures to exclude random or careless responses and other problems related to the lack of experimental control in the MTurk environment.

Pre-screening. We included data quality check items in the pre-screening questionnaire, which allowed us to identify attempts of careless or automatic responses. One of the items paraphrased an MEQ question (”How difficult do you find it to get up in the morning (when you are not awakened unexpectedly)?” in the original version as opposed to “How easy do you find it to get up…” in the modified item). The other item duplicated an MEQ question (”At approximately what time-of-day do you usually feel your best?") in a form of a text entry question, i.e., participants were required to manually type in the time-of-day in AM or PM format. Each of the questions of the pre-screening survey was presented on a separate page, and participants could not go back to check their previous responses. Second, we relied on the duration of the response to the pre-screening questionnaire as an indicator for random or careless responses. We considered response duration below 2 minutes and 30 seconds to be an indicator or careless or random responding for a survey containing 37-items. MTurkers who produced incorrect responses to the validity items or short response duration were not invited for participation.

Main experiment. Further data quality check items were included in the main part of the study. First, one of the MEQ items from the pre-screening questionnaire and three demographics questions were duplicated in the main experiment, allowing us to check for consistency of participant's response to these items across the two parts of the study. Second, we included the Instructional Manipulation Check (IMC; Oppenheimer, Meyvis, & Davidenko, 2009) in the concluding phase of the main part of the study. The IMC includes a question in a form of a block of text with lure responses, overall mimicking a typical multiple choice question. However, the long text block contains the instruction to submit a counterintuitive response. The accurate response to IMC could serve as one of the indicators that participants have been following the study instructions carefully. Finally, at the end of the experiment, participants had to answer two simple questions related to the content of the stimulus film. Specifically, participants were asked about the item that was stolen in the incident (multiple choice) and were instructed to indicate what the thief did with the stolen item (free entry field). Thus, one of the control questions referred to the middle of the stimulus film, while the other one concerned the detail at the end of the film.

This allowed us to detect careless or inattentive MTurkers and was also helpful in identifying cases when participants did not encode the stimulus fully due to some technical problems. Participant's response was counted as reliable only if they passed the IMC, showed consistent responses to the control items and responded to the questions related to the stimulus film accurately.

Possible character misrepresentation. Another potential problem is the possibility of character misrepresentation in Amazon MTurk samples. More specifically, experiments that target specific populations may encourage some MTurkers to falsify their identities (i.e., claim to belong to certain categories of population) or make multiple attempts to pass the screener questions to qualify for the main study and receive higher reimbursement. To address this concern, we adopted a two-step pre-screening procedure to prevent MTurkers from falsifying the answers to the pre-screening questionnaire to qualify for the main experiment (Wessling, Huber, & Netzer, 2017). Specifically, we aimed to create an impression that the pre-screening questionnaire was actually an independent survey focusing on the way sleep habits affect eating and caffeine consumption behaviour. There was no indication of a link between the pre-screening and the main study. The main study was visible only to MTurkers who were eligible and passed the initial data quality checks.

Participants. We continued data collection until we achieved the planned number of reliable responses. For this purpose, we pre-screened a total of 4,270 MTurkers. Among the pre-screened participants, 1,478 were morning-types and 568 were evening-types. A total of 365 proceeded to participate in the main study, of which 39 were excluded because they did not meet some of the data quality checks.

Upon completion of data collection, we discovered that the question asking participants to indicate their lineup decision appeared, due to an unanticipated technical error, on a separate rather than the same page as the lineup itself for 12 of the participants. In the event that this formatting discrepancy produced any distortion in responses provided by these participants, we collected data for a further 12 participants[4].

Hence, the final sample consisted of 324 participants (160 male, 163 female, 1 unspecified; age 19 to 66, *M* = 35.6, *Mdn* = 34 years). A total of 118 of them showed evening preference (36.4%, M_{MEQ} = 35.8, SD_{MEQ} = 3.99), and 206 participants were morning types (63.6%, M_{rMEQ} = 64.03, SD_{rMEQ} = 4.59). Participants received a $3 honorarium on completion of the experiment.

Design. The study used a two-factorial between-subjects design with time-of-day optimality (optimal versus non-optimal) and target presence (present versus absent) as independent variables. Participants were randomly assigned to be tested either at their optimal or the non-optimal time of day. We counterbalanced target presence of each lineup across the optimality conditions. Additionally, we partially counterbalanced the combinations of target presence in the three lineups: Participants received either one target-present and two target-absent lineups, or one target-absent and two target-present lineups.

In the thief lineups, we additionally manipulated mug shot exposure bias (bias versus no bias). In the biased condition, one individual appeared among both the mugshots and the lineup foils. Only one participant in the biased conditions selected the innocent suspect from the lineup, suggesting that the mug shot manipulation did not bias lineup decisions. Therefore, this element of the design will not be discussed further.

---

[4] Subsequently, we ran all analyses twice, once including the additionally collected 12 responses and once excluding them. The pattern of results was analogous. Therefore, in Results section we report statistical output for analyses performed including data collected from all participants.

Participants made three identification decisions. For each of the three lineups, target-presence was counterbalanced across the two optimality conditions. Identification accuracy (accurate versus inaccurate), post-decision confidence (on an 11-point scale ranging from 0-100%) and decision times served as dependent variables.

Materials.

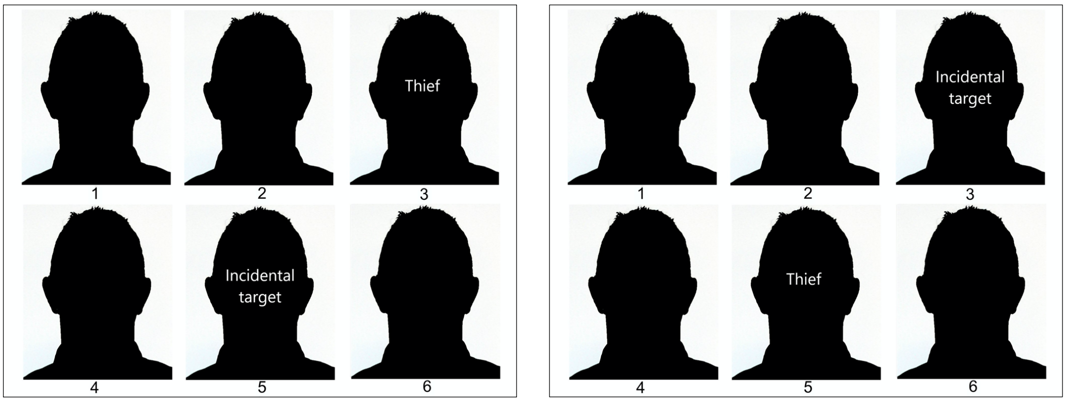
Pre-screening questionnaire. We used the full version of the MEQ (Horne & Östberg, 1976) to identify participants' time-of-day preference. The 19 original MEQ items were intermixed with filler questions related to participants' sleep, food and caffeine consumption habits to mask the aim of the pre-screening survey and provide additional support for the cover story. Example items are "How much control do you have over your eating between supper and bedtime?", "At what time do you prefer to drink your first caffeine-containing beverage?", "When you get up in the middle of the night, how often do you snack?" Additionally, the questionnaire included the two data quality check items.

Stimulus film. We used Film 2 from Experiment 1 as the stimulus event.

Mug books. We constructed two mug books, containing 16 head-and-shoulder photographs of white males that met the description of the thief from the stimulus film. The biased mug book contained a photograph of the innocent suspect; this person would also appear in the thief lineup as the incidental target. The unbiased mug book included a photograph of a control foil.

Photo lineups. We created six-person target-present and target-absent simultaneous lineups for each of the three targets appearing in the stimulus film, namely, the thief, the victim and the bystander. Target-absent lineups were constructed containing photographs of fillers that matched the description of the target in a similar way as in Experiment 1. For target-present lineups, we created six versions of lineups, with all possible combinations of the target with five of the foils from the target-present lineup. The target positions in the thief, victim, and bystander lineup were 3, 6, and 3, respectively. The incidental target (i.e., the innocent suspect who appeared among the biased mug shots) took position 5 in the thief target-present (see Figure 3, left panel) and target-absent lineups. Additionally, we created six extra target-present thief lineups in which positions of the thief and the innocent suspect were interchanged, that is, the thief took position 5 and the innocent suspect position 3 (Figure 3, right panel). Positions of the foils were randomized for each of the lineup versions. This resulted in a total of 12 versions of thief-present lineups, six versions of victim and six versions of bystander lineups.

Figure 3. Positions of thief and incidental target in target-present biased perpetrator lineups.



A pilot study was conducted to establish effective lineup sizes. For each of the targets, we tested the target-absent lineup and one randomly selected version of the target-present lineup. Tredoux's *E*s ranged from 3.5 to 4.6 (19 to 20 mock witnesses, total $N = 119$). The proportions of identification of the targets in the pilot study ranged from 5% to 30% (7.8% in the perpetrator lineup), with a critical ratio for difference from chance ranging between -2.4 and 1.3 (-2.01 in the perpetrator lineup).

Procedure

Participants who fulfilled the pre-screening criteria received access to the main study and a personal message inviting them for participation. As a cover story, we informed them that the experiment focused on the long-term effects of eating, caffeine-consumption habits and circadian rhythms on memory performance. We instructed participants not to consume alcohol for 8 hours prior to testing, more than two cups of coffee on the day of testing and sleep a minimum of six hours in the night prior to testing to exclude the possibility of confounding effects on alertness and memory performance.

Testing sessions took place either between 7:30 AM and 9:00 AM or between 8:30 PM and 10:00 PM. We used more extreme early morning and late evening hours compared to Experiment 1 to maximize the possible synchrony effect. Participants watched a stimulus film and were instructed to watch the film closely and pay attention to every detail and that they will be asked to act as eyewitnesses. Then, they provided answers to seven multiple choice questions concerning their food consumption habits (e.g., “How hungry are you usually in the morning?”); these were included for additional support of the cover story. Next, participants viewed the mug book. When all 16 mug shots had been presented, participants could make a selection or press the “Not present” option. As filler tasks, participants engaged in stem completion tasks adopted from Jacoby (1998) and an object search filler task for about 20 minutes. The results concerning the effect of testing optimality on performance in the stem-completion tasks are reported elsewhere.

Following the filler tasks, participants viewed the three lineups in succession and attempted to identify the thief, the victim and the bystander who were involved in the stimulus event. The thief lineup always appeared first. The remaining two lineups were presented in a random order. They were informed that the targets may or may not be present in the lineups and were encouraged to select the “Not present” option if they were not sure or didn’t know. After each of the decisions, participants indicated their confidence. Decision times for each identification decision were recorded automatically.

At the end of the experiment, participants were presented with a block data quality check items. The debriefing occurred upon completion of data collection.

Results

Data analyses. The pre-registered analysis plan was based on the assumption that presenting a mug shot of an innocent suspect in the biased conditions would affect the further identification decision in lineups. Because this manipulation was unsuccessful, our analyses deviated from the pre-registered analysis plan in two respects. First, we could not analyse the effects of our predictors on innocent suspect misidentifications. Second, we did not include the factor bias in our analyses of identification accuracy in the perpetrator lineups (i.e., we treated all identification decisions as unbiased) and hence conducted one analysis across all lineup decisions.

Identification performance. To test the effects of time-of-day optimality and target-presence on the likelihood that participants made an accurate identification decision, we entered time-of-day optimality, target-presence, and their two-way interaction in a GEE model. We deleted non-significant terms stepwise. The final model only included the main effect of target presence: Participants had higher odds of making an accurate decision from target-absent compared to target-present lineups, Wald $\chi2(1) = 9.08$, Exp(B) = 1.48, $p = .003$. The effect of testing optimality did not materialize[5]. Frequencies and proportions of different lineup outcomes for each of the lineups can be found in Table 3.

Table 3. Frequencies and proportions of different lineup outcomes at the optimal and the non-optimal time of day in Experiment 2.

Lineup	**Optimal**		**Non-optimal**	
	Frequency	Proportion	Frequency	Proportion
Thief				
Target-present				
Targets choices	21	23%	11	15%
Foil choices	20	22%	16	22%
False rejections	51	85%	47	64%
Total	92	100%	74	100%
Target-absent				
Correct rejections	60	68%	44	63%
Foil choices	28	32%	26	37%
Total	88	100%	70	100%
Victim				
Target-present				
Target choices	55	59%	44	63%
Foil choices	20	21%	14	20%
False rejections	19	20%	12	17%
Total	94	100%	70	100%

---

[5] Analyses of identification performance in solely perpetrator lineups showed no effect of time-of-day optimality.

Target-absent				
Correct rejections	54	63%	32	43%
Foil choices	32	37%	42	57%
Total	86	100%	74	100%
Bystander				
Target-present				
Target choices	47	55%	41	55%
Foil choices	20	23%	17	23%
False rejections	19	22%	16	22%
Total	86	100%	74	100%
Target-absent				
Correct rejections	39	41%	27	39%
Foil choices	55	59%	43	61%
Total	94	100%	70	100%

Therefore, we did not replicate the unusual findings obtained in Experiment 1, where performance in target-present lineups was better at non-optimal compared to the optimal time of day. We also did not observe the synchrony effect previously reported in recognition performance of non-facial stimuli (Intons-Peterson, Rocchi, West, McLellan & Hackney, 1999; May et al., 1993).

Differential effect of testing optimality in owls and larks. We ran exploratory analysis to test the possibility that the effect of testing optimality on identification accuracy manifested itself differently in participants with morning and evening chronotype. We entered chronotype (morningness versus eveningness), test time (morning vs evening), target-presence (target-present versus target-absent), their two-way and three-way interactions as predictors into the model. The final model included the main effects of target presence, Wald $\chi2(1) = 8.90$, Exp(B) = 1.48, $p = .003$, and chronotype, Wald $\chi2(1) = 10.39$, Exp(B) = 1.517, $p = .001$. This means that participants had higher odds of making an accurate decision from target-absent than target-present lineups and participants with evening preference were more likely to make an accurate identification decision than those with morning preference. The expected chronotype x test time interaction was not significant, indicating that the effect of chronotype occurred irrespective of the time of the day when testing occurred.

Sensitivity and response bias. Paralleling the signal detection analyses performed on data in Experiment 1, discriminability did not significantly differ between optimality conditions, with $d'_{\text{Non-Opt}} = 1.18$ versus $d'_{\text{Opt}} = 1.41$, $G = 0.66$, $p =$ .0.94. The response bias measures were again comparable for the two conditions, $c_{\text{Non-Opt}} = 0.76$; $c_{\text{Opt}} = 0.79$.

Figure 4 displays the ROC curves for optimal versus non-optimal testing conditions, with the diagonal line representing chance performance. The two curves nearly overlap, confirming that there was no benefit from matching participant's time-of-day preference at any of the confidence cut-off levels.

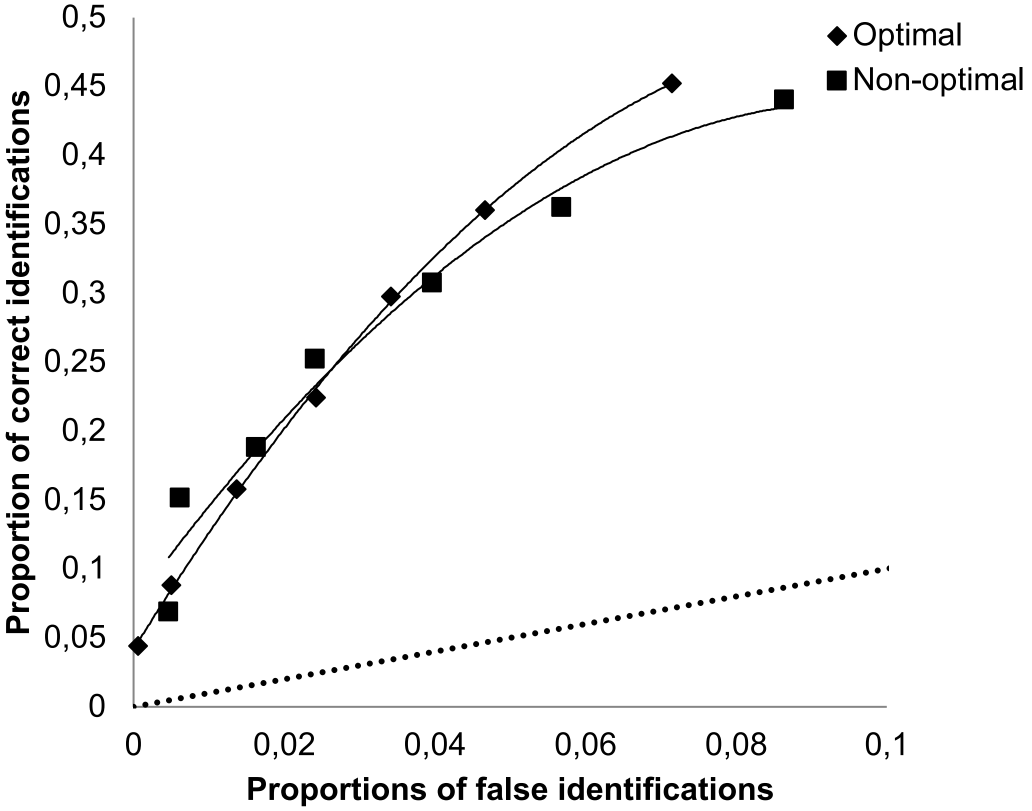
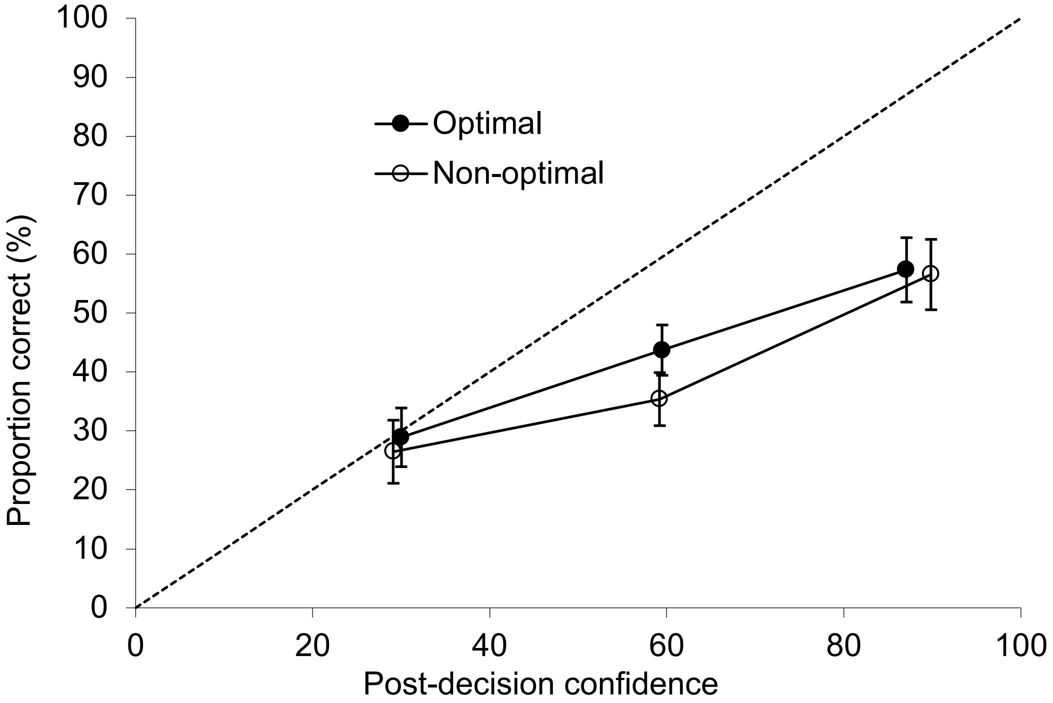


Figure 4. Receiver operating characteristics plots for the optimal and the non-optimal testing sessions across all lineups in Experiment 2.

To summarize, we observed no benefit for eyewitness identification decisions of matching testing time to participants' circadian preference. The finding that evening-type participants performed significantly better than morning-type participants may be attributable to differences in information processing styles between the two chronotypes. Evening time-of-day preference have been previously linked to superior performance in holistic information processing and appear to be better at processing non-verbal and emotional stimuli (Fabbri, Antonietti, Giorgetti, Tonetti, & Natale, 2007), that is, factors that are known to be relevant for face processing (e.g., Haxby, Hoffman & Gobbini, 2002; Tanaka & Farah, 1993).

Effect of time-of-day optimality on the confidence-accuracy relationship. To test the effect of time-of-day optimality on the confidence-accuracy relationship, we entered identification accuracy (accurate versus inaccurate), choosing (selection versus rejection), and time-of-day optimality (optimal versus non-optimal), as well as their two- and three-way interactions as predictors in the initial GEE model. The final model contained a significant Accuracy x Choosing x Optimality interaction, Wald $\chi^2(1) = 5.47$, $p = .019$. Separate analyses split by time-of-day optimality revealed a significant simple Accuracy x Choosing interaction at the non-optimal, Wald $\chi^2(1) = 16.91$, $p < .001$, but not the optimal time of day, Wald $\chi^2(1) = 0.69$, $p = .406$. Further analyses performed on choosers and nonchoosers that were tested at the non-optimal time of day showed that accurate choosers were more confident than inaccurate choosers, $b = 1.31$, Wald $\chi^2(1) = 33.07$, $p < .001$. This effect was not statistically significant for nonchoosers, $b = 0.97$, Wald $\chi^2(1) = 0.26$, $p = .610$. Contrary to the findings of Experiment 1, these results show that time-of-day optimality did affect the postdictive value of confidence. However, the direction of the effect was opposite than we hypothesized: Confidence in choosers was more predictive of accuracy at non-optimal compared to the optimal time of day.

2.2.3.1. Calibration of confidence measures. The confidence-accuracy correlation was significant for choosers, *r*(548) = .26, *p* < .001, and for nonchoosers, *r*(420) = .113, *p* = .020. Figure 5 (top panel) displays the confidence-accuracy calibration curves for choosers at the optimal (number of observations *n* = 300) and non-optimal (*n* = 250) time of day. Visual inspection of the curves reveals no significant differences between optimal and non-optimal testing (all the confidence intervals overlap). Similarly, nonchoosers did not show any significant differences between the two conditions: The calibration curves for both optimal (*n* = 240) and non-optimal (*n* = 182) sessions almost parallel the X-axis (see lower panel of Figure 5). In line with findings obtained in Experiment 1, the confidence intervals for all calibration statistics overlap (see Table 4).



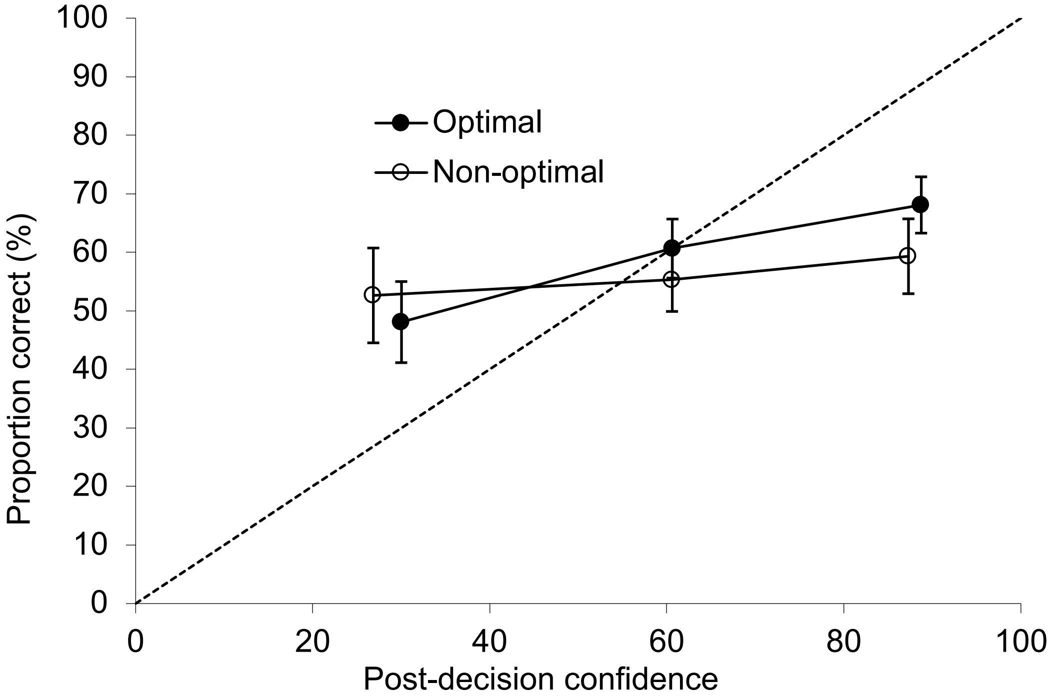


Figure 5. Post-decision confidence-identification accuracy calibration curves (and 95% CI bars) for the optimal and the non-optimal time-of-day in choosers (top panel) and nonchoosers (lower panel) in Experiment 2.

Table 4. Calibration measures for optimal and non-optimal testing sessions split by choosers and nonchoosers

	Experiment 2			
	Optimal		Non-Optimal	
		95% CI		95% CI
Choosers				
C	.011	.000; .023	.023	.004; .041
O/U	.055	.000; .111	.116	.055; .176
NRI	.067	.004; .129	.102	.022; .181
Nonchoosers				
C	.029	.008; .051	.031	.005; .057
O/U	-.058	-.123; .006	-.008	-.086; .069
NRI	.021	-.016; .059	.004	-.014; .022

Effect of time-of-day optimality on the decision time-accuracy relationship. To establish the effect of time-of-day optimality on the decision-time-accuracy relationship, we entered identification accuracy (accurate versus inaccurate), choosing (selection versus rejection), and time-of-day optimality (optimal versus non-optimal) in a GEE model. As in Experiment 1, we log-transformed decision times (log base 10) due to significant positive skewness and kurtosis. The final model contained only the main effect of Choosing, $b = .05$, Wald $\chi2(1) = 9.23$, $p < .001$. Participants who made a selection were faster than those who rejected the lineup.

Discussion

Over the decades, research has collected sufficient evidence showing that that time of day can affect cognitive performance across an array of cognitive domains, which also include memory. This guided our interest in investigating whether time of day is also a factor of relevance for the eyewitness memory field, in particular with regards to eyewitness identification performance. We expected that testing participants at the optimal as opposed to the non-optimal time of day would have beneficial effect on identification accuracy.

Results of the current study showed that the optimal testing had no beneficial effect on identification accuracy. These findings do not parallel previously reported patterns in recognition memory, which is normally superior during circadian peaks as opposed to circadian troughs (Intons-Peterson et al., 1999; May et al., 1993; Petros et al., 1990).We also did not replicate the unusual 'reverse synchrony' effect in eyewitness identification performance found in Experiment 1.

Several explanations may account for this discrepancy. Previous studies that found synchrony effect in recognition memory tested verbal and non-facial pictorial memory rather than memory for faces. Could face recognition performance be immune to "classical" circadian patterns found in other types of recognition memory? This would not be the first instance of such dissociation: Processing of faces is traditionally considered to be a highly specialized function that is relatively independent from other types of information processing, and examples of selective impairments in one of the memory domains with others remaining intact are well-known (Robotham & Starrfelt, 2017).

Alternatively, the effect size of time-of-day optimality on face recognition performance could in reality be smaller than we anticipated based on prior research that used non-facial stimuli. Hence, our failure to detect the synchrony effect should not necessarily be interpreted as evidence of absence of circadian variations in face recognition performance per se. The face recognition paradigm with multiple recognition trials could offer more statistical power to detect such an effect, if present, compared to the eyewitness memory paradigm with only few identification decisions.

Our additional interest was in whether the non-optimal time of day magnifies unwanted influence of factors that are known to have a biasing effect on eyewitness identifications. To explore this possibility, we exposed some of the participants to a mug shot of an innocent suspect prior to administering the lineup. This manipulation has previously been used to generate biased outcomes in subsequent identification in a scenario when a lineup includes the photograph of the aforementioned innocent suspect (Deffenbacher, Bornstein & Penrod, 2006). Our interest was in whether non-optimal testing would increase this biasing effect. Unfortunately, the mug shot manipulation did not appear to be successful. The failure to replicate mug shot exposure effect in Experiment 2 can be linked to a short retention interval between exposure to mug shots and a subsequent lineup identification, which was about 20 minutes in our experiment as opposed to a minimum of 48 hours and up to a week in prior research (e.g., Brown, Deffenbacher, & Sturgill, 1977; Goodsell, Neuschatz, & Gronlund, 2009; Memon, Hope, Bartlett, & Bull, 2002; Perfect & Harris, 2003). Anticipating the difficulties of conducting multi-session experiments in MTurk environment, we refrained from using an experimental design with a long retention interval between mug shot exposure and identification decisions. Future research employing a successful biasing manipulation can investigate the possibility that non-optimal testing may magnify the detrimental effect of the factors that are known to reduce identification accuracy.

We were also interested in whether testing optimality would affect the postdictive value of confidence and decision times. Based on the optimality hypothesis (Deffenbacher, 1980), we expected a stronger confidence-accuracy relationship in choosers at the optimal compared to the non-optimal time of day. Results showed that an opposite effect: the postdictive value of confidence was stronger in participants that were tested at circadian troughs compared to circadian peaks. A similar statistical model in Experiment 1 did not reveal such an effect, i.e. confidence-accuracy relationship was not moderated by testing optimality. In the meantime, a similar 'reverse synchrony' effect was observed calibration analyses in Experiment 1: Participants were significantly better calibrated in some of the confidence categories at non-optimal compared to the optimal time of day.

These data provide further support for theory-driven confidence judgments hypothesis as a potential explanation of the way time-of-day optimality affects confidence-accuracy relationship (Palmer, Brewer, Weber, & Nagesh, 2013). This hypothesis states that, if aware of the factor that negatively affects memory, participants may take this information into account and adjust their metamemory judgements appropriately. This might have been the case in the current experiment; that is, our participants may have been aware of the possible cognitive impairments due to the circadian troughs and taken this information into account when indicating their post-decision confidence.

It is important to take into account that the findings of the two experiments point in the same direction but the source of evidence is not consistent. In Experiment 1, some support for theory-driven confidence judgements hypothesis came from the calibration analyses, while in the current experiment it was evident from the GEE analyses. The fact that evidence in favour of theory-judgment confidence judgment hypothesis comes from different statistical approaches is possibly explainable by the differences in the experimental designs of the two studies (within- versus between-subjects) and / or differences in the encoding and retrieval stimuli (e.g., differences in the lineup construction approaches). The robustness of the obtained effect may be confirmed in future research by testing whether similar tendencies are observable across a variety of experimental stimuli.

In Experiment 1, we found that decision times were predictive of accuracy only in optimal sessions. We were interested in replicating this finding in Experiment 2. However, our results showed that decision times had no postdictive value in the current experiment: We did not observe a commonly found effect where accurate choosers make faster identification decisions than inaccurate choosers (Brewer & Wells, 2006; Dunning & Stern, 1994; Sporer et.al., 1995; Sauerland & Sporer, 2007, 2009). This may serve as an indication that assessing the predictive value of decision times in MTurk studies can be problematic due reduced control over the experimental environment.

It is important to note that our findings are limited to situations when all other factors that are known to affect memory performance are optimal. In both of the experiments conducted, the stimulus film presentation occurred at the beginning of the testing session, eliminating possible fatigue effects. Participants were asked to watch the stimulus film closely and warned that they would later be asked to act as eyewitnesses. Each of the targets' faces was visible for a sufficient amount of time (e.g., Memon, Hope & Bull, 2003) to ensure robust encoding strength. All identification instructions were consistent with good practice in the eyewitness identification procedures (Steblay, 1997). In real-life scenarios, however, encoding conditions are not always optimal, nor are the identification procedures always best examples of good police practices. Additionally, certain categories of eyewitnesses may be especially prone to memory errors; for instance, this concerns the elderly participants due to the age decline in memory performance (see Fitzgerald & Price, 2015 for a recent meta-analysis). It remains unclear whether non-optimal testing would have an additive detrimental effect under such conditions.

Ensuring best retrieval conditions for eyewitnesses is critical, which necessitates a close examination of factors that contribute to memory performance in eyewitnesses. We tested whether time-of-day optimality could be another factor of high relevance to the eyewitness memory field. For instance, is it possible to increase identification accuracy of eyewitnesses by administering the lineup at peak hour of the day? Based on results of Experiment 1 and Experiment 2, the takeaway message for the policymakers is straightforward: We found no evidence supporting the idea that daily variations in performance affect eyewitness identification performance. In other words, with respect to time-of-day optimality, neither the time at which the event was witnessed nor the timing of the lineup administration appears to affect identification outcome.

Chapter 4: Circadian Effects on Face Recognition and Source Memory Performance

Abstract

The circadian rhythm regulates arousal and activity levels throughout the day and determines hours of optimal cognitive performance. However, circadian fluctuations in face recognition performance received little attention so far in the research literature. The current experiment investigated the effects of time-of-day optimality on the ability to recognize faces and discriminate between the contexts in which where they were encountered. We tested 91 morning- and evening-type participants either at their optimal or the non-optimal time of day. Participants encoded some faces in a crime-related context, whereas other faces were presented in a neutral context. We expected that testing during circadian peaks as opposed to circadian troughs would result in better recognition performance. We also hypothesised that, when asked to recognise faces from the crime-related context, participants would be better at excluding familiar but irrelevant faces from the neutral context at the optimal compared to the non-optimal time of day. Results showed that overall recognition performance was not affected by testing optimality. When asked to exclude familiar but irrelevant faces, participants showed only slight benefit from non-optimal testing that was not statistically significant. Future research is necessary to explore the underlying mechanisms of time-of-day effects in face recognition and source memory performance in vulnerable witnesses.

Keywords: circadian rhythm, synchrony effect, face recognition, source memory, source monitoring

Introduction

Have you ever recognized a person but were unable to recall where you knew them from? This relatively common experience reveals an important fact about the way our memory works: Familiarity is not always accompanied by the correct recollection of the memory source (Johnson, Hashtroudi & Lindsay, 1993). Whereas in everyday life the consequences of such memory source errors are mostly benign, in legal contexts the costs can be more severe. In some cases, an innocent suspect may become familiar to an eyewitness from mug shot viewings, repeated identification procedures, or simply because they were a bystander at the crime scene (Deffenbacher, Bornstein, and Penrod, 2006). A failure to attribute familiarity to the correct source can result of misidentifications and wrongful convictions (Lindsay, 2007). Therefore, factors that increase the likelihood such memory errors require careful examination. Our body clock could be one such factor: It is known that certain hours of the day are associated with a decline in cognitive performance caused by the circadian rhythm (Schmidt, Collette, Cajochen, & Peigneux, 2007). This observation motivated us to investigate whether testing at non-optimal times of day could be among the risk factors for errors in source memory for faces.

Our body clock, also known as the circadian rhythm, coordinates numerous aspects of our physiology, cognition and behaviour (Halberg et al., 2003; Refinetti, 1999). Among other things, it determines optimal times for engaging in various activities and modulates cognitive performance levels across the day: We generally perform better at circadian peaks and worse at circadian troughs (Schmidt et al., 2007). The exact timing of circadian phases can vary across individuals, resulting in inter-individual differences in parts of the day that are optimal (Adan et al., 2012; Horne & Östberg, 1976). The so-called morning types perform best early in the day, whereas evening types peak in their performance in the evening hours (Goldstein et al. 2007; Levandovski, Sasso, & Hidalgo, 2013). This pattern is known as the synchrony effect: Performance is a function of synchrony, or alignment of the time when certain task is performed with the ongoing circadian phase (e.g., May, 1999; May & Hasher, 1998; Nowack & Van Der Meer, 2018).

In terms of episodic memory performance, the pattern of circadian variations depends on the nature of the memory test. In traditional tests such as recall and recognition, performance is generally better at the optimal time of day (May et al., 1993; Petros et al., 1990; Ryan et al., 2002). In these tests, participants are explicitly instructed to use the studied material to complete the task. Some other manifestations of memory are quite different: They may occur outside of our awareness and rely less on intention and control. For instance, in a phenomenon known as priming, performance on a task is facilitated by prior stimulus encoding. Priming effects are not guided on intention or depend on conscious effort; rather, they occur automatically and outside our awareness (Bargh & Chartrand, 2000).

Such unconscious memory effects appear to be affected by the circadian rhythm a different manner compared to standard recall and recognition tasks. For instance, May, Hasher and Foong (2005) report dissociations in the way the circadian rhythm affects priming as opposed to a cued recall task performance. When explicitly instructed to use the studied words to complete the stems, participants showed a classic synchrony effect pattern: They performed better at circadian peaks compared to circadian troughs. Surprisingly, the time-of-day optimality effect on priming was reversed: participants were more likely to unintentionally use earlier-encoded items on seemingly unrelated category generation and stem completion tasks at non-optimal compared to the optimal time of day.

Similar findings were reported by Yang, Hasher, & Wilson (2007) using the so-called speeded retrieval procedure. First, participants were trained to rapidly retrieve words from semantic memory in a stem-completion task. Unbeknown to participants, the procedure went on to turn into a memory test, in which some of the stems could be completed words that had been presented to participants earlier. The change from automatic to controlled processing was operationalized in terms of increase in the retrieval speed, the assumption being that controlled retrieval requires more time than automatic retrieval. Results showed synchrony effect only in controlled retrieval (participants who slowed down showed better performance at the optimal compared to the non-optimal time of day), while automatic manifestations of memory remained unaffected by the circadian phase.

Finally, this pattern was confirmed by Puttaert, Adam, and Peigneux (2018), who used the process dissociation procedure (the PDP) to measure relative contribution of automatic and controlled processes to memory performance at the optimal and the non-optimal time of day. The PDP is a method that allows to measure relative contributions of automatic and controlled processes to memory performance (Jacoby, 1991). In line with predictions of dual process model of memory performance, Puttaert et al. (2018) found estimates of that controlled processing were higher at the optimal compared to the non-optimal time of day, whereas controlled processing was unaffected by time of testing[6].

Therefore, it appears that the circadian rhythm affects our conscious, intentional uses of memory in such a way that we perform better when tested at our optimal time of day (e.g., May et al., 1993; Petros et al., 1990). In the meantime, more

---

[6] It should be noted though that this experiment differed from traditional synchrony effect research in was the timing of optimal and non-optimal sessions was determined. Participants reported their own individual 2-hour period of optimal and non-optimal performance; then participants were randomly assigned to be tested either at their (self-reported) optimal or the non-optimal time of day.

automatic influences of memory seem to be either unaffected by testing optimality or show reverse synchrony, in which automatic influences of memory are stronger at the non-optimal time of day (May et al., 2005; Yang et al., 2007). These findings parallel dissociations commonly found with other factors that affect performance, such as divided attention or age-related decline in memory (e.g., Jacoby, 1998; Yonelinas & Jacoby, 2012; Yonelinas, 2002).

Dual-process models of memory serve to provide explanatory accounts of this dissociation (for a review, see Smith & DeCoster, 2000). According to these models, two fundamentally different categories of cognitive processes can contribute to memory performance. The so-called automatic processes operate outside of conscious awareness and voluntary control. They are inevitably triggered by the presence of the respective stimuli in the environment and tend to be executed fully. Controlled processes, on the contrary, are voluntary and intentional. They are conscious and require effort in order to be executed (Bargh, 1994).

Automatic processes are generally believed to require little to no cognitive resources for their execution, and, therefore, can run in parallel with other cognitive processes with no decline in efficiency. They are unaffected by many factors that are generally impair cognition, including stress or arousal (Hasher & Zacks, 1979, 1984). On the contrary, controlled cognitive processes rely on available cognitive resources for their execution. Performance on tasks that rely on these processes is impaired under the conditions when cognitive resources are limited (Smith & DeCoster, 2000). This also appears to be the case with time-of-day optimality: Reduced cognitive capacity at circadian troughs impairs performance on tasks with a stronger controlled component (e.g., May et al., 1993; Petros et al., 1990; Ryan et al., 2002), whereas automatic memory influences remain unaffected (May et al., 2005; Yang, Hasher, & Wilson, 2007).

Both automatic and controlled processes contribute to a certain extent to performance on any cognitive task (Jacoby, 1991; Yonelinas & Jacoby, 2012). For instance, in dual-process models of recognition, two memory processes can contribute independently to recognition judgments. One process is known as familiarity: Items that were encountered before seem more familiar compared to items that have not been encountered. Familiarity may serve as basis for recognition judgments, that is, we can use familiarity as criterion for saying that we encountered a certain person before. Recollection is a different process that involves retrieval of some qualitative information about the studied event, which can also serve as basis for recognition judgments (Atkinson & Juola, 1974; Jacoby & Dallas 1981; Mandler, 1980; for a review of dual-process theories of recognition, see Yonelinas, 2002).

Familiarity appears to be a relatively automatic process: It is automatically evoked by the presence of the stimulus in the environment, is faster and more effortless, and appears to rely less on intention. On the contrary, recollection is a controlled, intentional process. Its efficiency highly relies on cognitive capacities available at its execution (Jacoby, 1991).

Both familiarity and recollection contribute to face recognition processes in eyewitnesses. If presented with the culprit during identification, the eyewitness may be able to engage in recollection and retrieve some qualitative information about the event (e.g., “I remember this man: he was the one who was holding the gun”). In case recollection fails, the culprit may still be more likely to be selected based merely on familiarity from having been encountered by the witness during crime (”I think have seen this person before - he may have been the man from the crime scene”).

However, let us consider a different scenario in which the suspect presented during identification, though innocent, incidentally happens to be familiar to the eyewitness. The suspect may have been present at the crime scene as a bystander, or their face may have become familiar from previously administered identification procedures or mug shot viewings (Deffenbacher et al., 2006). In these scenarios, a response based solely on familiarity could result in a misidentification. Instead, an accurate identification decision requires recollection of the context in which the suspect was encountered before (”They already showed me a mug shot of this man two weeks ago”).

Memory for the context in which information was encoded is often referred to as source memory (Johnson et al., 1993). Generally, source memory tasks are more difficult compared to standard recognition tasks, and impairing factors can show disproportionate deficits in source memory performance compared to simple recognition (for a review, see Mitchell & Johnson, 2009). The dual-process approach maps well onto this pattern: Impairing factors, that often affect recollection but leave familiarity intact, would produce larger impairments in source memory compared to recognition memory (Yonelinas, 1999).

In the current experiment, we investigated whether time-of-day optimality impairs general face recognition performance as well as memory for the context in which the face was encountered. We recruited morning- and evening-type participants and tested them either at their optimal or the non-optimal time of day. Two sets of face stimuli were presented at encoding; some faces were embedded in a crime-related context, whereas other faces were presented in a neutral context. On a recognition test, participants simply had to indicate whether they encountered the face earlier in the experiment. We expected participants’ recognition memory to be better in participants that were tested at the optimal time of day as opposed to non-optimally tested participants. Specifically, we expected that participants would produce higher hit and lower false alarm rates in the optimal compared to the non-optimal sessions.

To measure source memory, we administered another test in which we asked participants identify faces exclusively from the crime-related context. On this task, a yes response to the face from a neutral context could occur if the face appears familiar but participant failed to retrieve the context in which the face was encountered. We hypothesised that participants would fail to exclude a higher proportion of irrelevant faces at circadian troughs compared to circadian peaks. We also hypothesized that the effect of optimality would be more pronounced in the source memory task compared to the recognition task, as source memory relies more on recollection compared to recognition memory task.

Experiment 3

Method

Participants. To determine the required sample size, we conducted a-priori power analysis for a one-tailed *t* test with G*Power v3.1 (Faul, Erdfelder, Buchner, & Lang, 2009; Faul, Erdfelder, Lang, & Buchner, 2007). Based on the results reported by May et al. (2005), we expected an effect size $d = 0.52$. We used an alpha error probability of .05 and a power of .80, which resulted in a required sample size of 94.

A total of 346 individuals who expressed their interest in participation were pre-screened for their chronotype using the Morningness-Eveningness Questionnaire (MEQ; Horne & Östberg, 1976). Based on results of the pre-screening, morning- and evening-type individuals were identified and invited for participation in the experiment. We managed to test 91 of the planned 94 participants (25 male, 66 female; age 18 to 54, $M = 22.13$, $Mdn = 21$ years), resulting in a power of .79 to detect an effect of the expected size. The sample included university students (*n* = 87) and members of the general population (*n* = 4). Among them, 57.1% were evening types (*n* = 52, M_{MEQ} = 34.37, SD_{MEQ} = 5.05) and 42.9% were morning types (*n* = 39, M_{MEQ} = 63.77, SD_{MEQ} = 4.46). We only included Caucasian participants to avoid cross-racial bias (Meissner & Brigham, 2001; Wilson, Berstein, & Hugenberg, 2016).

Design. The experiment used a three-factorial mixed design, with instruction type (recognition test vs source memory test) and context type (criminal vs neutral) serving as within-subjects factors and optimality (optimal vs non-optimal) as a between-subjects factor. Participants were randomly assigned to be tested either at their optimal or the non-optimal time of day.

During the study, they encoded two sets of face stimuli. One of the sets was accompanied with a scenario that put the presented faces into a crime-related context (i.e. "…You will now be presented with the faces of the hooligans"), whereas the other set was accompanied with a neutral-context scenario (i.e. "You will now see the people you've encountered in the supermarket while shopping"). The two face stimuli sets (Set 1 vs Set 2) were counterbalanced across the contexts (criminal vs neutral) in such a way that half of participants encoded Set 1 in criminal context and Set 2 in the neutral context, while the other half encoded Set 2 in the criminal context and Set 1 in the neutral context. These two combinations of sets of faces across contexts were counterbalanced across the optimality conditions.

Every participant underwent two memory tests. One test was a standard face recognition task, in which participants had to indicate whether they saw the face earlier in the experiment, irrespective of the context. The other test was a source memory task, in which participants had to exclude faces from the neutral context and only say yes to faces from the criminal context. In each task, participants were presented with old (earlier-studied) and distractor (non-studied) faces.

Half of the earlier-studied faces in each task were from the crime-related context and the other half were from the neutral context. We divided each of the encoded face sets into two subsets that were rotated through the two memory tests, counterbalanced across the optimality conditions. Another half of the faces presented for memory tests were new (non-presented) faces. For this purpose, we used to sets of distractor faces that were rotated throughout the two tests and counterbalanced across the optimality conditions. The proportion of yes responses to faces from each context and to distractor faces served as the outcome variables.

Materials.

Pre-screening questionnaire. The questionnaire from Experiment 2 was used to pre-screen participants for their chronotype. The questionnaire contained MEQ items (Horne & Östberg, 1976) and filler questions about food and caffeine consumption habits and sleep.

Face stimuli. We used face stimuli from the Karolinska Directed Emotional Faces database[7] (Lundqvist, Flykt, & Öhman, 1998) for the current experiment. The database contains photographs of 70 male and female amateur actors displaying different emotional expressions (in the current experiment, we only used neutral-expression versions). All the actors were between 20 and 30 years old and were wearing the same grey T-shirt. When the stimuli were prepared, the actors were not permitted to wear earrings or eyeglasses, visible makeup, beards or moustaches. Importantly, for each of the actors, the database includes two versions of each facial expression taken at two separate occasions. Having two different photographs of the

---

[7] The database is freely available for scientific purposes and is accessible online via www.KDEF.se.

same person available allowed us to use different photographs of the same individual at encoding and at test, which is a preferable way of testing face recognition performance (Burton, 2013).

For our experiment, we selected photographs of 64 faces with a neutral emotional expression. A total of 32 faces (16 male, 16 female) were used as targets. They were further divided into two sets 16 faces each (8 male, 8 female) that were rotated through the context (crime-related vs neutral). Faces were presented against a background image depicting the respective context (see Figure 1). For each context, we created four versions of each face set with the order of faces randomized to control for order effects in memory performance. In each encoding context, participants were presented with one of the four versions that was selected randomly.

For the memory tests, each subset was further divided into two subsets consisting of eight faces (4 male, 4 female) that were rotated through the test conditions (recognition task vs source memory task). We also selected 32 faces that were used as filler faces in the memory tests. They were divided into two sets 16 faces each (8 male, 8 female) and were rotated through the two tasks (recognition task vs source memory task). The full counterbalancing table can be referred to in Appendix B.

Figure 6. Example of encoding stimuli presented in crime-related context (top row) and neutral context (bottom row).

Procedure. Participants were recruited via advertisements at Maastricht University campus. They were told that the experiment concerned differences in food and caffeine consumption habits between morning- and evening-type individuals and their effect on memory performance. We instructed participants not to consume alcohol 18 hours prior to testing, avoid caffeine and chocolate on the day of testing and schedule participation only if they slept a minimum of six hours in the night prior to testing. Morning sessions took place between 7:40 AM and 9:00 AM, and evening sessions were scheduled between 8:30 and 9:30 PM.

First, participants encoded the face stimuli in the criminal context. We asked participants to imagine the following scenario: They are waiting for a bus late at night. There has been an important football derby in the town and a large group of football hooligans is behaving violently nearby. The hooligans approach a single fan of the other team and start insulting him. Very quickly the encounter becomes violent: the hooligans start pushing and beating him. The participant decides to call the police. These instructions were accompanied by a context-cue: a dark street with some police cars and a policeman. After that, we informed participants that they would be presented with the faces of the hooligans. We instructed them to pay attention as they may be asked to identify them later. Participants saw 16 faces, presented one at a time for one second followed by a 0.5-second interval. We presented the faces against the background that accompanied the imagery instructions.

Next, participants engaged in a filler task (object search) for 5 minutes, after which they encoded the next set of faces in the neutral context. To introduce the context, we asked participants to imagine that they are going to the supermarket on a rainy day. At the entrance, they drop off the beloved umbrella they received as a gift from their grandmother. After finishing shopping, they realize that their umbrella is gone; instead, they notice another umbrella that looks very similar to theirs. Someone must have taken their umbrella by accident and left a different one instead. These instructions were accompanied with an image of the supermarket aisle. After that, we told participants that they would now see the people they encountered in the supermarket. We told participants to pay attention to the faces as one of them might have their umbrella. The faces for encoding were presented in the same way as in the criminal context (one second per face with a 0.5-second interval in between) against the background of the supermarket aisle (for stimuli examples, See Figure 1).

After the encoding phase, participants again engaged in a 5-minute filler task (object search) and answered 21 questions about their sleep, food and caffeine consumption habits in order to provide additional support to the cover story. Finally, participants were presented with two memory tests, a recognition test followed by a source memory test. In the recognition test, we reminded participants that they encoded pictures of individuals in two different contexts and asked them to indicate whether they saw each of the faces in either context. A total of 16 studied faces (8 faces from each of the two contexts) and 16 new (non-studied) faces were presented one at a time in random order. Participants were asked to indicate if they had seen the face earlier in the experiment by pressing the ‘yes’ key; if they had not encountered the face before, they were asked to press the ‘no’ key.

On the source memory test, we told participants that, unlike in the previous task, they should only select the faces that were shown in the criminal context (i.e. the hooligans). Similarly to recognition test, 16 old and 16 new faces appeared on the screen one at a time in random order. Participants pressed ‘yes’ if they had seen the face in the criminal context. If they had seen the face in the supermarket context or had not encountered the face before, they were instructed to press ‘no’. Participants had unlimited time to make their decisions.

Participants received 5-euro gift vouchers or participation credit for their participation. We debriefed participants upon the completion of the data collection.

Results

Recognition memory. One outlier produced zero hits and false alarms on the recognition task and zero false alarms on the source memory task, whereas the proportion of correct responses on the source memory task was high (.88). We ran all analyses twice, once including the data from this participant and once excluding it. The results did not significantly differ in either case. We further report the analyses and descriptives excluding data from the outlier.

We ran independent-samples t-tests to determine if there were differences in overall hit rates (proportion of yes responses to faces from both contexts) and false alarm rates between optimal and non-optimal sessions on the recognition task. There was no statistically significant difference in hit rates between optimal (M = .47, SD = 0.20) and non-optimal (M = .50, SD = 0.20) time of day, $t(88)$ = - 0.926, p = .505, d = 0.15. Analogously, false alarm rates did not signifcantly differ bewteen optimal (M = .22, SD = 0.16) and non-optimal sessions (M = .25, SD = 0.15), $t(88)$ = - 1.169, p = .722, d = 0.19. Sensitivity was d' = 0.81 at the optimal time of day and d' = 0.76 at the non-optimal time of day. Response bias was c = 0.96 in optimal and c = 0.74 in non-optimal sessions.

Source memory. We ran a three-way mixed ANOVA to understand whether optimality affected participants’ recognition performance on the face recognition task and ability to exclude irrelevant faces on the source memory task. Optimality (optimal vs non-optimal), instruction type (recognition test vs source memory test), context (crime-related vs neutral), their two- and three-way interactions were entered as predictors; proportions of ‘yes’ responses to faces studied in each context (i.e. independent of accuracy) served as the dependent variable.

The model revealed a significant main effect of context, *F*(2, 89) = 18.473, *p* < .001, partial η^2 = .172: Participants provided more yes responses to faces from criminal compared to non-criminal context (M_{crim} = .48 and $M_{non\text{-}crim}$ = .41). However, a simple instruction type x context interaction failed to reach significance, *F*(4, 89) = 0.623, *p* = .432, partial η^2 = .007. Instead, we observed a significant main effect of instruction: Participants provided more yes responses to old faces on the face recognition task (*M* = .49) compared to the source monitoring task (*M* = .40), irrespective of context in which they were studied, *F*(2, 89) = 10.307, *p* = .002, partial η^2 = .104.

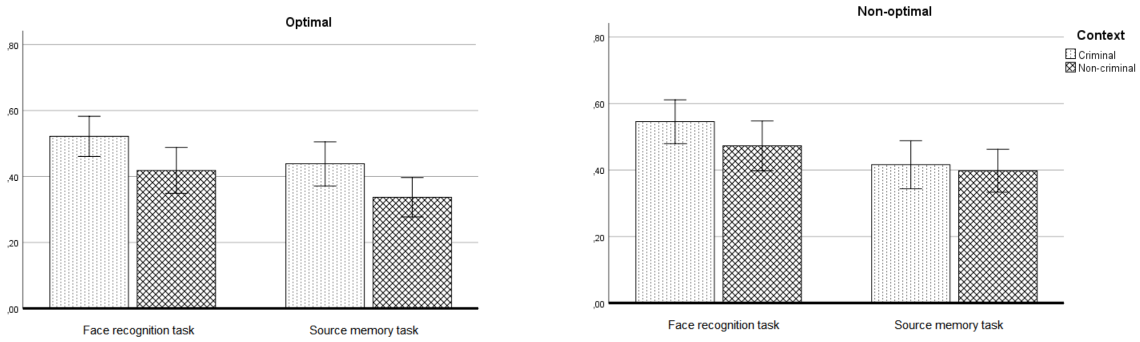


Figure 7. Proportion of Yes responses to faces encoded in each of the two contexts in the recognition task and the source memory task (and 95% CI error bars) at the optimal and the non-optimal time of day.

Numerically, participants on the source monitoring test provided lower rates of yes responses to faces from non-criminal compared to criminal context ($M_{non\text{-}crim}$ = .34 and M_{crim} = .44) at the optimal time of day, whereas this was not the case at the non-optimal time of day (M_{crim} = .42 and $M_{non\text{-}crim}$ = .40 respectively, see Figure 2). However, this difference was not reliable: the three-way optimality x instruction type x context interaction was not statistically significant, $F(6, 89) = 0.545$, $p = .462$, partial $\eta^2 = .006$. Mean proportions of yes responses across the experimental conditions are presented in Table 1.

Table 1

Mean proportions of Yes responses [and 95% CI] to faces from criminal and non-criminal contexts across the two instruction conditions at the optimal and the non-optimal time of day.

		Optimal	Non-optimal
Recognition task	Criminal	.52 [.46; .58]	.55 [.48; .61]
	Neutral	.42 [.35; .49]	.47 [.40; .55]
Source memory task	Criminal	.44 [.37; .51]	.42 [.34; .49]
	Neutral	.34 [.28; .40]	.40 [.33; .46]

Discussion

In our previous experiments, and contrary to what we expected, we failed to observe a common synchrony effect pattern, in which performance is better during circadian peaks as opposed to circadian troughs (May et al., 1993; Petros et al., 1990; Ryan et al., 2002). One possible explanation of this discrepancy pertained to the nature of the experimental paradigm used to examine the phenomenon. Specifically, eyewitness memory models differ in terms of the way stimuli are encoded and the memory test is administered and allow to collect only a limited number of recognition decisions per participant. Thus, in Experiments 1 and 2, participants encoded a stimulus film and made identification decisions from target-present and target-absent lineups.

In the current experiment, we used a face recognition paradigm that includes multiple recognition decisions collected per participant. This approach offered higher statistical power to detect possible circadian fluctuations in face recognition performance and was better aligned with common methodologies in the synchrony effect literature. However, the current results show no evidence in favor of hypothesis that the optimal testing leads to higher recognition performance: Participants tested at the time of day congruent with their circadian typology performed very similarly compared to non-optimally tested participants. This observation contrasts with findings reported in previous studies that used non-facial stimuli to examine synchrony effects in episodic memory performance (e.g., Anderson et al., 1991; Petros et al., 1990; May et al., 1993). However, these findings are consistent with the results we obtained using the eyewitness memory paradigm in Experiments 1 and 2.

Processing of faces is a highly specialized cognitive function, and performance in the domain of face recognition is dissociable from other types of recognition performance (Bruce & Young, 2012; Robotham & Starrfelt, 2017). It is possible that circadian variations in face recognition performance are not as pronounced as in other types of memory performance. Alternatively, the circadian rhythm in face recognition can be expressed in a different manner than predicted by the synchrony effect model, so that hours of the day that are optimal in terms of processing of faces differ from hours that are optimal for other types of memory performance. This possibility should be explored in future research using highly controlled paradigms such as the forced desynchrony protocol (Wright, Hull, Hughes, Ronda, & Czeisler, 2006).

Along with general face recognition performance, we were also interested in potential negative effect of non-optimal testing on source memory that is, the memory for the context in which a face was encountered. Eyewitnesses make source memory judgements if presented with a suspect that is innocent but nonetheless familiar to them from other (non-criminal) contexts. In this case, eyewitnesses are required to attribute the suspects' familiarity to the correct source, and the failure to do so may result in misidentification of an innocent person (e.g., Deffenbacher et al., 2006; Read, Tollestrup, Hammersley, McFadzen, & Christensen, 1990; Ross, Ceci, Dunning, & Toglia, 1994).

To experimentally recreate a of situation that requires source memory judgments, we administered a test in which we instructed participants to exclude faces from the neutral context and only say yes to a face if they remember encountering it in the criminal context. A yes response to the face from the neutral context on this test can occur if the face seems familiar but the memory for source is impaired. We expected that participants would be more likely to make such errors when tested at non-optimal compared to the optimal time of day. On a descriptive level, results showed that participants appeared to perform better at the optimal time of day, where proportion of yes responses to faces from the neutral context was lower compared to faces from the criminal context, compared to the non-optimal time of day, where the proportions of yes responses to faces from the two contexts were nearly identical. Nonethelss, this difference failed to reach statistical significance.

Participants provided an overall lower proportion of 'yes' responses in the source memory test compared to recognition test; that is, they were less prone to identifying faces as 'criminals'regardless of the context in which the faces were encoded. It appears that the task instructions induced an overall lower rate of 'yes' responses, resulting in lower rates of yes responses to lures at a cost of reduction in hit rates. These results show that participants found the source memory test difficult and adopted a conservative strategy when responding to faces from both contexts. Optimal testing produced only small non-significant beneficial effect on performance in this task.

The circadian rhythm is an important factor affecting performance across a large number of cognitive domains. Can face recognition and memory for the situations in which they were encountered can benefit from testing during circadian peak as opposed to circadian trough? In the current experiment, we did not find evidence for such circadian-related impairments in healthy young adults. Aditional research is necessary to test the possibility of the synchrony effect in older populations, which are known to have age-related decline in memory performance (Fitzgerald & Price, 2015) and are more prone to source memory errors (e.g., Benjamin & Craik, 2001; Schacter, Kaszniak, Kihlstrom, & Valdiserri, 1991; Trott, Friedman, Ritter, Fabiani, & Snodgrass, 1999). It is also important to further elucidate possible differences in the way the circadian rhythm affects memory for faces compared to other types of stimuli.

Chapter 5: General Discussion

Accurate and informative evidence provided by eyewitnesses can play a pivotal role in investigations in many criminal cases (Peterson, Sommers, Baskin, & Johnson, 2010; Ridley, Gabbert, & La Rooy, 2013; St-Yves, 2015). However, eyewitness testimony can be unreliable, resulting in counterproductive investigative actions, miscarriages of justice and dramatic failures to meet the global objectives of the law (Clark, Benjamin, Wixted, Mickes, & Gronlund, 2015; Forst, 2013; Wagenaar, 2009). The incidence of prosecution of innocent individuals resulting from erroneous eyewitness identification is alarmingly high (Innocence Project, 2019). These facts necessitate a close examination of factors that can potentially contribute to memory performance in eyewitnesses.

Basic laboratory research often marks directions for this investigation by identifying general principles of cognition. The applied methodologies used in the eyewitness memory literature, on the other hand, make it possible to study how the general principles of memory functioning manifest themselves in real-life eyewitnesses. As an example, consider the reconstruction principle of memory (Bartlett, 1932). This discovery offered theoretical background for uncovering the dangers of exposing eyewitnesses to post-event misinformation (Loftus, 2005; Loftus, Miller, & Burns, 1978), which contributed to development of evidence-based recommendations for investigative interviewing protocols (e.g., Gabbert, Hope & Confrey, 2018). The eyewitness memory field is rich in examples of basic laboratory discoveries being used to strengthen our understanding of the eyewitness memory phenomena (Wells, 1978).

Despite the significant progress that has been made (e.g., Wells, Memon, & Penrod, 2006; Wells & Olson, 2003), the relevance of other potential contributors to eyewitness memory performance remains unknown. The so-called synchrony effect in cognitive performance is among the most interesting factors that are yet to be explored in this context. It is now a general agreement in the psychological literature that the amount of cognitive capacities available to us does not stay constant across the day (Schmidt, Collette, Cajochen, & Peigneux, 2007). Driven by our body clock, our performance at certain hours of the day is better compared to others (Valdez, 2019). In the meantime, criminal activity takes place around the clock (e.g., Felson & Poulsen, 2003), and one may become a witness to a crime at any moment of the day. It is therefore inevitable that some eyewitnesses rely on their memory for events they saw during hours when their capacity to encode new information was poor. Furthermore, obtaining testimony during suboptimum performance periods could lead to further impairments in accuracy and informativeness of eyewitness statements and identifications.

In basic laboratory experiments, recall and recognition performance is consistently better at the optimal compared to the non-optimal hours (May, Hasher, & Foong, 2005; May, Hasher, & Stoltzfus, 1993; Petros, Beckwith, & Anderson; 1990; Puttaert, Adam, & Peigneux, 2018; Ryan, Hatfield, & Hofstetter, 2002; Yang, Hasher & Wilson, 2007; Yoon, 1997), suggesting potential relevance of time-of-day optimality to the eyewitness memory field. Nonetheless, this hypothesis had not previously been tested in a systematic matter using the eyewitness memory paradigm. The current research programme presents the first comprehensive attempt to empirically test possible time-of-day optimality effects across various aspects of eyewitness memory performance.

Overview of the Findings

In Experiment 1, we compared performance in three different domains of eyewitness memory at the optimal and the non-optimal times of day. After encoding the stimulus events, participants provided free narratives and answered cued questions about the incident and the people involved. Results showed no effect of testing optimality on accuracy of free reports or answers to cued questions. We also asked participants to identify people involved in the incidents. To our surprise, and contrary to previous synchrony effect findings in recognition memory (e.g., May et al., 1993), identification accuracy in target-present lineups was unexpectedly higher at the non-optimal time of day. Performance in target-absent lineups was not affected by testing optimality. Decision times were also predictive of accuracy at the non-optimal compared to the optimal time of day, in line with previous data on the effect of circadian arousal on long-term memory access speed (Anderson, Petros, Beckwith, Mitchell and Fritz, 1991). The postdictive value of confidence, however, showed some evidence of “reversed” synchrony: Highly confident choosers were significantly better calibrated at the non-optimal compared to the optimal times of day.

Finally, we compared false memory rates in the visual version of the DRM paradigm at the optimal and the non-optimal times of day. Results showed comparable false memory rates across the optimality conditions. This finding was in line with previously reported results on production of false memories in young adults (Intons-Peterson, Rocchi, West, McLellan & Hackney, 1999).

The superior identification performance at the non-optimal time of day in Experiment 1 was very surprising: Prior research predicted the opposite effect, that is, better performance at circadian peaks. It was also surprising that this increase in hit rates was not accompanied by an increase in false-alarm rates, which is a typical pattern in recognition studies (Wixted & Mickes, 2014). We were also surprised that the effect was specific to target-present lineups. In addition, the size of the observed effect was small. These factors guided our interest in testing whether the obtained effect would replicate.

For this purpose, we conducted Experiment 2. Enhancements in the experimental design allowed us to fully eliminate possible expectancy effects, potential effects of lineup composition, and increase generalizability of our findings to a broader population. In addition, in Experiment 2 we aimed to test the idea that non-optimal testing could magnify proneness to memory biases. For this purpose, we introduced bias into participants' identification decisions by exposing them to a mug shot of an innocent suspect prior to lineup identification. The photograph of the innocent suspect also appeared in the lineups, and we expected participants to erroneously identify him based on familiarity gained during the mug shot exposure (Deffenbacher, Bornstein, & Penrod, 2006). We expected this negative tendency to be magnified by non-optimal testing.

Results showed that mug shot exposure did not produce a biasing effect we had anticipated, making it impossible to assess whether testing optimality had a moderating potential for this negative, and potentially dangerous, effect commonly reported in the literature[8]. In terms of identification accuracy, we found no evidence of significant differences between optimal and non-optimal sessions. Inferential analyses showed that confidence was predictive of accuracy at non-optimal but not optimal times of day, paralleling evidence from calibration analysis in Experiment 1.

Strikingly, previously reported patterns of superior performance at the optimal compared to the non-optimal time of day were not replicated in either of the two experiments we conducted. What are the possible reasons for the discrepancy of our

---

[8] We addressed this issue by including a more effective biasing manipulation in Experiment 3, as discussed below.

results with a seemingly the "classic" synchrony effect patterns reported in the literature? One possibility is that the circadian arousal affects face recognition performance in a different manner than recognition of other types of stimuli. Differential impairments in face processing as opposed to processing of other types of stimuli are well-acknowledged (Bruce & Young, 2012; Robotham & Starrfelt, 2017), and our results suggest that the circadian rhythms is yet another factor that affects face processing in an atypical manner.

In order to further explore this possibility, we used a modification of the face recognition paradigm in Experiment 3. Participants encoded sets of faces and were presented with a recognition test at either optimal or the non-optimal time of day. We expected recognition performance to be better at circadian peaks as opposed to circadian troughs. Multiple recognition trials allowed us to collect a significantly larger number of data points per participant compared to Experiments 1 and 2, offering statistical power to detect even small-size effects. Contrary to our hypothesis, results showed no evidence of superior recognition performance at the optimal compared to the non-optimal time of day.

Our second aim in Experiment 3 was to test whether memory for contexts in which faces were encountered is better at circadian arousal peaks compared to arousal troughs. The use of the face recognition paradigm allowed us to design a more potent biasing manipulation compared to mug shot exposure we relied upon in Experiment 2. Specifically, we embedded face stimuli into two distinct contextual scenarios. Some faces were presented within a context of a criminal scenario, whereas other faces were presented within a neutral context. We then administered a memory test that required participants to discriminate between the two different contexts in which the faces were encoded. Specifically, we asked participants to recognize faces they had seen in the crime-related context, ignoring the faces from the neutral context. We expected that familiarity gained during encoding of neutral-context faces would misguide participants to incorrectly recognize some of neutral-context faces as ‘criminals’. Our main interest was in testing whether these errors would be more likely to occur in non-optimal compared to optimal sessions. Contrary to our hypothesis (but consistent with overall pattern of findings our line of research), participants showed no evidence of significant benefit in source monitoring task from optimal testing.

To summarize, across a battery of memory tests included in the three experiments within the current research programme, we consistently found no evidence of superior performance at circadian peaks as opposed to circadian troughs. In addition, identification accuracy in Experiment 1 showed a small but significant effect in the opposite direction than predicted by the synchrony effect model: Performance was counterintuitively better at the non-optimal time of day. This pattern of findings is contrary to our expectations regarding the effect of time-of-day optimality on memory performance in eyewitnesses.

Interpreting Results

Task difficulty. Factors that affect cognitive performance can produce differential levels of impairment across different levels of task difficulty. Tasks that require high cognitive resources for their execution and are generally affected by impairing factors to a large extent, whereas tasks that require less effort may be affected less or not affected at all (e.g., Yonelinas & Jacoby, 2012). The memory tests we have administered in our experiments are not homogeneous in terms of their difficulty, and we expected that non-optimal testing would result in stronger impairments in more difficult tasks compared to easier tests.

The face recognition task in Experiment 3, as well as identification tasks in Experiments 1 and 2 belong to the category of recognition tests. They are generally believed to have lower levels of difficulty compared to other types of memory tests. Recognition tasks provide the strongest retrieval cues in the form of a representation of the to-be-recognized item. Hence, an accurate response in a recognition test can be based solely upon the feeling that the item seems familiar. Familiarity is an automatic process that requires little to no cognitive resources, which can make it immune to many factors that negatively affect our cognition (Jacoby, 1991). Troughs in circadian arousal are no exception: They appear to have no impairing effect on automatic processes (May et al., 2005; Yang et al., 2007). Therefore, the absence of the synchrony effect in recognition tests in our experiments, though contrasting previous findings by May et al. (1993), is less surprising and has parallels in other known memory-impairing effects in cognitive psychology[9].

The source memory task employed in Experiment 2 has a higher level of difficulty compared to a standard recognition test (Mitchell & Johnson, 2009). In this task, participants had to discriminate between faces encountered in two different contexts. In order to complete the task, it was necessary to go beyond familiarity-based responses and retrieve qualitative information about the context in which each face was encountered. This process is known as recollection and it differs from familiarity in many important aspects (e.g., Jacoby, 1991). Unlike familiarity, recollection is reliant on available cognitive resources and therefore can be strongly affected by factors that reduce our attentional capacities (Yonelinas, 1999). Recent data shows that the dips in cognitive performance caused by the circadian rhythm act

---

[9] Consider, for instance, such a robust and reliable factor as age-related decline in memory (Fitzgerald & Price, 2015). Whereas it is strongly pronounced in recall tasks, it can show no impairments on recognition (e.g., Schonfield & Robertson, 1966).

in a similar manner: They appear to affect recollection while leaving automatic memory processes intact (Puttaert et al., 2018).

These findings offer interesting predictions regarding performance on tasks in which familiarity and recollection oppose each other. When a familiarity-based response is incorrect and needs to be ignored based on information retrieved in recollection, suboptimal processing conditions increase the chance of memory biases (for a review, see Yonelinas, 2001). The source memory task in Experiment 3 was designed specifically to create such an opposition between recollection and familiarity. All the studied faces could appear familiar to participants, but some of them belonged to irrelevant context and thus needed to be rejected. We expected that participants would be better at ignoring familiar but irrelevant faces at the optimal time of day, when their cognitive capacities are better. To our surprise, participants in this task showed only minor benefit from optimal testing, which was not statistically significant. This finding is not in line with predictions of the synchrony effect model.

Finally, free and cued recall are the most difficult memory tasks we used in our experiments. They provide the lowest amount of retrieval cues and therefore primarily rely upon recollection. Time-of-day optimality has been consistently shown to affect performance on these tasks (e.g., Intons-Peterson et al., 1999; Petros et al., 1990; Puttaert et al., 2018). These patterns are in line with the cognitive capacity explanation of the synchrony effects in cognitive performance: Recollection is an effortful process largely reliant upon cognitive capacities for successful execution (e.g., Schoenfield & Stones, 1979) and therefore should show most robust impairments during circadian troughs. Nonetheless, free narratives and answers to cued questions in Experiment 1 were not subject to this standard synchrony effect pattern. This finding is highly surprising: We expected that the synchrony effect in eyewitness memory, if present, should be easily detectable in these tests highly reliant upon availability of cognitive capacities.

To summarize, the fact that we did not observe impairing effects of non-optimal testing across a wide battery of memory tests of varying difficulty levels is surprising from the perspective of the dual process theory of memory performance (Yonelinas, 1999) and contrasts previously reported synchrony effect patterns in laboratory tasks of analogous levels of task difficulty (Schmidt et al., 2007). To interpret these discrepancies, we turn our focus to methodological differences between prior research and our experimental protocols.

Encoding instructions. Similar to the protocol used in the previous synchrony effect research (e.g., May et al., 2005; May et al., 1993; Petros et al., 1990), participants in our experiments were instructed to play close attention to the presented stimuli and warned about the forthcoming memory test. Nonetheless, certain aspects of the eyewitness memory paradigm may have put our participants in a more beneficial position compared with participants in previously conducted synchrony effect studies.

In Experiments 1 and 2, we informed participants before presenting the stimulus event that they would be asked to serve as eyewitnesses. Compared with a generic warning that participants' memory for the presented stimuli will be tested, this information gives participants more detailed understanding about the nature of the subsequent memory test. Based on common knowledge about eyewitness testimony, our mock witnesses 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, etc. Therefore, it is possible that prior knowledge on 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.

An additional benefit of informing participants about the fact that they would act as eyewitnesses concerns the activation of participants' schemata (i.e. general knowledge, beliefs expectations) of crime. Processing the to-be-encoded stimuli in relation to pre-existing schemata generally has beneficial effects on encoding, including higher accuracy rates (e.g., Hastie & Kumar, 1979; for a review, see van Kesteren et al., 2012). Combined, these factors may have resulted in stronger encoding of the presented stimuli compared to previously conducted research showing synchrony effects in memory performance.

Retrieval instructions. The retrieval instructions in the eyewitness memory paradigm also differ from instructions normally used in previous synchrony effect studies. The idea that one should aim to perform well on a memory test is certainly inherent in any memory test. However, we specifically emphasized the importance of providing as complete and accurate answers as possible. We also explicitly discouraged participants from guessing details they could not remember. Consistent with standard methodologies in eyewitness memory research, identification instructions clearly specified that the perpetrator may or may not be present in the lineup, and we encouraged participants to select the "Not present" option if they were not sure or did not know. Such instructions are known to induce a more conservative response strategy and encourage a neutral position towards the presence of the target in the lineup (Steblay, 1997). No equivalent of such instructions was present in the protocols of previous studies that showed synchrony effects in memory performance (Schmidt et al., 2007). These aspects of our protocols may have affected retrieval strategies in our participants in a different manner than in basic memory tests used previously to demonstrate synchrony effects in memory performance.

Stimuli Type. Previous research showing synchrony effect patterns in performance tested memory for encoded prose passages, sentences, word lists, word stems and pictorial stimuli (e.g., Intons-Peterson et al., 1999; May et al., 1993; Puttaert et al., 2018). By contrast, participants in Experiments 1 and 2 encoded stimulus events depicting staged crimes and subsequently identified individuals involved in the incidents. In Experiment 3, we used the face recognition paradigm. The fact that we did not observe a standard synchrony effect pattern in tasks involving recognition of faces may suggest that the circadian rhythm affects processing of faces differently than memory for other types of stimuli.

Face processing is a highly specialized function that differs from other types of recognition (Haxby et al., 2001; Kanwisher, McDermott, & Chun, 1997). Impairments in the face recognition domain are known to be dissociable from impairments in recognition of non-facial stimuli (Bruce & Young, 2012; Robotham & Starrfelt, 2017). Our findings may indicate that circadian-related impariments could be another example of such dissociations.

To summarize, the analysis of methodological differences between previously conducted synchrony effect research and our studies reveals important differences in task type, stimulus type, and encoding strategies that are likely to account for the discrepancy between our findings and previously reported data. Notably, all of the discussed factors are specific to the eyewitness memory paradigms and reflect the differences between conditions to which eyewitnesses are exposed in real-life situations as opposed to participants in a standard laboratory experiments. Our results highlight theoretical gaps in understanding the mechanisms behind time-of-day fluctuations in memory performance and raise the issue of constraints of generality of standard laboratory research to applied settings, such as eyewitness memory contexts.

Theoretical Contribution

The synchrony effect and compensatory resource allocation mechanisms.

Troughs in cognitive performance are often explained in terms of decreased availability of cognitive resources at certain points of the day (Hirst & Kalmar, 1987; Necka, 1997). Recent research investigated allocation of cognitive resources at the optimal and the non-optimal hours of the day by measuring participants' pupil dilation as they were performing a semantic analogy task (Nowack & Van Der Meer, 2018). This study demonstrated that participants were able to use different resource allocation strategies to partially compensate for circadian-related impairments. These findings suggest that the dropdown in performance at the non-optimal time of day can be mediated by the type of resource allocation strategies: Participants who are able to avoid wasteful allocation of limited attentional capacities during circadian troughs show lower levels of performance impairments (Nowack & Van Der Meer, 2018).

The absence of drops in free and cued recall performance in Experiment 1 can be linked to the fact that participants were able to efficiently allocate limited attentional resources. Compared with prior research that relied upon generic memory tests, our experimental protocols provided participants the opportunity to prioritize the limited attentional capacities in non-optimal sessions efficiently, allowing them to offset the potential impairing effects of circadian troughs. The possibility that the impairing effect of suboptimal factors can be partially compensated by retrieval instructions is an interesting direction for future research. Future experiments can investigate whether the impairing effect of non-optimal testing on performance of mock witnesses materializes when participants are not warned about the subsequent memory test. It is also important to understand whether the negative effects of suboptimal retrieval and identification instructions (Steblay, 1997) can be more strongly pronounced at non-optimal hours.

Identifying the strategies we can use to counteract the limited availability of cognitive resources offers another interesting avenue for future research. In particular, future studies can use the eyewitness memory paradigm in combination with pupil dilation techniques to measure allocation of cognitive resources in eyewitness memory. It would be interesting to identify the differences in compensatory mechanisms mock witnesses use to offset effects of non-optimal testing. Similar methodology could be used to identify whether encoding and retrieval instructions can encourage the use of efficient compensatory resource allocation strategies when our cognitive capacities are at their low point.

The circadian rhythms and metamemory. Our results also outline an interesting direction of future research into time-of-day optimality effects on postdictors of identification accuracy. A traditional approach to studying the way suboptimal factors affect our metamemory judgments is rooted in the optimality hypothesis proposed by Deffenbacher (1980). The general idea behind this theory is that suboptimal factors weaken our metamemory judgments in a similar manner as they impair our memory performance. It follows from this that the predictive value of confidence should be lower in the presence of suboptimal factors at encoding and / or retrieval.

In the meantime, an increasing body of evidence suggests that this may not always be the case. In their theory-driven confidence judgments hypothesis, Palmer, Brewer, Weber, and Nagesh (2013) suggest that we can take into account the presence of impairing factors and adjust our metamemory judgments appropriately. The effects we observed in Experiments 1 and 2 appear to offer support for this idea. In Experiment 1, highly confident choosers were significantly better calibrated in their confidence judgments at the non-optimal compared to the optimal time of day. Experiment 2 showed that overall confidence-accuracy relationship was stronger (i.e. confidence in choosers was more predictive of accuracy) in non-optimal compared to optimal sessions.

The way eyewitnesses perceive troughs in their performance and other factors potentially damaging to their memory in relation to the actual impairment caused by these factors deserves more attention in the eyewitness memory literature. It would be interesting to combine these research questions within a single research line exploring interrelations between the extent to which we are aware of the presence of potentially impairing factors, the compensatory mechanisms we employ to counteract these impairments, and the way this is reflected in our metamemory judgments.

Face recognition and fluctuations in arousal. The current research programme is novel in showing that face recognition performance may not follow the standard synchrony effect patterns. This appears to be a consistent finding in our studies, as evident from identification decisions (Experiments 1 and 2) as well as face recognition data (Experiment 3). To the best of our knowledge, time-of-day effects in memory performance have not been previously tested using facial stimuli.

From a cognitive perspective, processing of faces differs from processing of other types of information (Schwartz, 2014). Faces are processed holistically (e.g., Richler, & Gauthier, 2014) and are distinct in terms of allocation of attention during encoding. For instance, when presented with two faces simultaneously, encoding of one of the faces requires us to suppress processing of features of the other face, whereas this is not the case when we are simultaneously presented with two objects (e.g., Bindemann, Jenkins, & Burton 2007; Boutet & Chaudhuri, 2001; Palermo & Rhodes, 2002). Considering the fact that circadian variations in cognitive performance are generally construed in terms of availability of attentional capacities (Valdez, 2019) and abilities of efficient allocation of cognitive resources (e.g., Nowack & Van Der Meer, 2018), these peculiarities of face processing may account for the fact that the circadian performance cycles are expressed differently in face recognition performance.

The possibility that attentional mechanisms underlying face processing may offer extra gain in terms of compensating for circadian troughs in attentional resources outlines an interesting avenue for future research. It is unclear whether face recognition is fully immune to circadian fluctuations in arousal or less sensitive to them. Future studies can investigate this possibility by using an experimental design that includes multiple measurements throughout the day comparing recognition of objects as opposed to faces.

From a neuroscientific perspective, circadian rhythms are not simply a product of oscillations generated in suprachiasmatic nuclei. Areas of our brain responsible for memory functioning appear to show their own autonomous oscillations that can contribute to daily cycles in memory functioning (Snider, Sullivan, & Obrietan, 2018). In this regard, it may be important to consider that face recognition relies upon specialized areas of the brain different from areas involved in other types of recognition memory. It remains unknown whether the peripheral oscillators in these specialized areas function in a different manner. If this turns out to be the case, face recognition may show a divergent pattern of daily fluctuations in performance compared to other types of stimuli. The combination of brain imaging techniques combined with the strengths of the forced desynchrony protocol may offer promising discoveries about dissociations in circadian fluctuations in face recognition performance and recognition of objects and verbal stimuli.

Practical Implications

We designed the current research program to answer the applied question whether circadian arousal could be among the factors affecting eyewitness memory performance. Across multiple tests, we did not find any empirical evidence supporting the idea that eyewitness memory performance is affected by testing optimality in a similar manner as in other domains of memory performance. This is true for such aspects of eyewitness memory as informativeness and accuracy of free narratives and answers to cued questions, as well as identification performance and ability to discriminate between contexts in which faces were encountered. Therefore, we can conclude that circadian troughs in cognitive performance do not lead to significant reduction of the evidential value of eyewitness testimony in healthy young adults.

It is important to emphasize that our findings are limited to situations of optimal encoding conditions and retrieval instructions. Real-life eyewitnesses may encode events under less favourable conditions, such as insufficient lighting (Wagenaar & van der Schrier, 1996), suboptimal distance (Lindsay, Semmler, Weber, Brewer, & Lindsay, 2008), or short exposure duration (Memon, Hope, & Bull, 2003). Moreover, retrieval conditions such as substandard identification instructions (Steblay, 1997) or poor lineup construction (Fitzgerald, Price, Oriet, & Charman, 2013) increase the likelihood of memory errors. Some of these factors may be more pronounced at circadian arousal troughs, which outlines numerous perspectives for future research. Additionally, elderly eyewitnesses are generally poorer eyewitnesses (see Fitzgerald & Price, 2015 for a recent meta-analysis). 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.

Our findings also suggest that postdictive value of decision times can be eliminated at the non-optimal time of day. These findings be of interest to memory researchers who rely on decision times as their outcome measure, as they appear to suggest that that time of the day when the test is administered may confound the decision-time-based outcomes.

Limitations

The current study shares its limitations with other studies that rely on the classical synchrony effect paradigm. For instance, the design of our study does not allow us to isolate the effect of circadian rhythm from that of the homeostatic sleep pressure, that is, the decrease in arousal levels associated with the increased amount of time spent awake (Van Dongen & Dinges, 2003). A demanding paradigm known as the forced desynchrony protocol can allow researchers to overcome this limitation. In this protocol, participants are placed in an environment that isolates them from the external time givers, such as light and social timing cues. This allows to manipulate the sleep and wakefulness cycle in such a way that the duration of the “day” is other than 24 hours (e.g., 19 hours or 28 hours). As a result, the sleep-wake cycle and the circadian rhythm become desynchronized, allowing researchers to disentangle their complex interactions and measure the effects of each of them separately (Wright, Hull, Hughes, Ronda, & Czeisler, 2006).

However, these protocols are logistically complicated, considered extreme, and test participants under highly artificial conditions, which undermines the applicability of the findings to the real situations to which eyewitnesses are exposed. Keeping in mind limited resources and the applied focus of the research question, we were not in a position to consider using the forced desynchrony protocol in the current programme of research. Nonetheless, this protocol can serve as an excellent methodological platform for disentangling complex interactions of circadian and sleep-related effects on face recognition performance.

Our study only tested morning- and evening-type individuals, which raises the issue of constraints of generality. Further research into the specifics of time-of-day variations in identification performance in intermediate types is necessary. Comparing memory performance at the optimal and the non-optimal hours in intermediate-type eyewitnesses is important for understanding the variability in performance patterns in the general population. Another limitation of our experiments concerns reliance on self-report tools in determining participants' periods of optimal performance. Future studies may enhance the precision of classifying participants into different chronotype groups by collecting additional physiological measurements, such as cortisol levels and body temperature (Blatter & Cajochen, 2007).

Finally, encoding and retrieval in all our experiments took place in the same experimental session. This design does 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, that is, by employing a testing optimality (optimal versus non-optimal) x memory stage (encoding versus retrieval) design.

References

Adan, A., & Almirall, H. (1991). Horne & Östberg morningness-eveningness questionnaire: A reduced scale. *Personality and Individual Differences*, *12*, 241–253. doi: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*, 1153-1175. doi:10.3109/07420528.2012.719971

Anderson, M., Petros, T., Beckwith, B., Mitchell, W., & Fritz, S. (1991). Individual differences in the effect of time of day on long-term memory access. *The American Journal of Psychology*, *104*, 241–255. doi:10.2307/1423157

Andrade, M. M., Benedito-Silva, A. A., Menna-Barreto, L. (1992). Correlations between morningness-eveningness character, sleep habits and temperature rhythm in adolescents. *Brazilian Journal of Medical and Biological Research*, *25*, 835-839.

Antonenko, D., Diekelmann, S., Olsen, C., Born, J., & Mölle, M. (2013). Napping to renew learning capacity: Eenhanced encoding after stimulation of sleep slow oscillations. *European Journal of Neuroscience*, *37*, 1142-1151. doi:10.1111/ejn.12118

Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology: I. Learning, memory and thinking.* W. H. Freeman

Baehr, E. K., Revelle, W., Eastman, C. I. (2000). Individual differences in the phase and amplitude of the human circadian temperature rhythm: With an emphasis on morningness-eveningness. *Journal of Sleep Research, 9,*117-27. doi:10.1046/j.1365-2869.2000.00196.x

Bailey, S.L., & Heitkemper, M. M. (2001). Circadian rhythmicity of cortisol and body temperature: Morningness-eveningness effects. *Chronobiology International*, *18*, 249-261. doi:10.1081/CBI-100103189

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Handbook of social cognition: Basic processes; Applications* (p. 1–40). Lawrence Erlbaum Associates, Inc

Bargh, J. A., & Chartrand, T. L. (2000). The mind in the middle: A practical guide to priming and automaticity research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (p. 253–285). Cambridge University Press

Bartlett, F. C. (1932). Remembering: A study in experimental and social psychology. Cambridge University Press.

Benjamin, A. S., & Craik, F. I. (2001). Parallel effects of aging and time pressure on memory for source: Evidence from the spacing effect. *Memory & Cognition*, *29*, 691-697. doi:10.3758/BF03200471

Bindemann, M., Jenkins, R., & Burton, A. M. (2007). A bottleneck in face identification: Repetition priming from flanker images. *Experimental Psychology*, *54*, 192-201. doi:10.1027/1618-3169.54.3.192

Blatter, K., & Cajochen, C. (2007). Circadian rhythms in cognitive performance: Methodological constraints, protocols, theoretical underpinnings. *Physiology & Behavior*, *90*, 196–208. doi:10.1016/j.physbeh.2006.09.009

Bodenhausen, G. (1990). Stereotypes as judgmental heuristics: Evidence of circadian variations in discrimination. *Psychological Science*, *1*, 319–322. doi:10.1111/j.1467-9280.1990.tb00226.x

Borbély, A., & Achermann, P. (1999). Sleep homeostasis and models of sleep regulation. *Journal of Biological Rhythms*, *14*, 559–570. doi:10.1177/074873099129000894

Borbély, A., Daan, S., Wirz-Justice, A., & Deboer, T. (2016). The two-process model of sleep regulation: A reappraisal. *Journal of Sleep Research*, *25*, 131–143. doi:10.1111/jsr.12371

Bousfield, W. A., & Cohen, B. H. (1953). The effects of reinforcement on the occurrence of clustering in the recall of randomly arranged associates. *The Journal of Psychology*, *36*, 67–81. doi:10.1080/00223980.1953.9712878

Boutet, I., & Chaudhuri, A. (2001). Multistability of overlapped face stimuli is dependent upon orientation. *Perception*, *30*, 743-753. doi:10.1068/p3183

Brackmann, N., Sauerland, M., & Otgaar, H. (2018). *Developmental trends in lineup performance: Adolescents are more prone to innocent bystander misidentifications than children and adults.* Manuscript submitted for publication.

Bradfield, A., & Wells, G. (2000). The perceived validity of eyewitness identification testimony: A test of the five Biggers criteria. *Law and Human Behavior*, *24*, 581–594. doi:10.1023/a:1005523129437

Brewer, N., & Wells, G. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied*, *12*, 11–30. doi:10.1037/1076-898x.12.1.11

Brewer, N., & Wells, G. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied*, *12*, 11-30. doi:10.1037/1076-898x.12.1.11

Brewer, N., Vagadia, A. N., Hope, L., & Gabbert, F. (2018). Interviewing witnesses: Eliciting coarse-grain information. *Law and Human Behavior, 42*, 458–471. doi:10.1037/lhb0000294

Brown, E., Deffenbacher, K., & Sturgill, W. (1977). Memory for faces and the circumstances of encounter. *Journal of Applied Psychology, 62*, 311–318. doi:10.1037/0021-9010.62.3.311

Bruce, V., & Young, A. (2012). *Face perception.* New York, NY: Psychology Press. doi:10.4324/9780203721254

Buhrmester, M., Kwang, T., & Gosling, S. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*, 3-5. doi:10.1177/1745691610393980

Burton, A.M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *Quarterly Journal of Experimental Psychology*, *66*, 1467-1485. doi:10.1080/17470218.2013.800125

Buysse, D., Reynolds, C., Monk, T., Berman, S., & Kupfer, D. (1989). The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Research*, *28*, 193–213. doi:10.1016/0165-1781(89)90047-4

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*, 240–245. doi:10.1016/j.sleep.2008.01.007

Cann, D., McRae, K., & Katz, A. (2011). False recall in the Deese–Roediger–McDermott paradigm: The roles of gist and associative strength. *Quarterly Journal of Experimental Psychology*, *64*, 1515–1542. doi:10.1080/17470218.2011.560272

Cariou, M., Galy, E., & Mélan, C. (2008). Differential 24-hour variation of alertness and subjective tension in process controllers: Investigation of the relationship with body temperature and heart rate. *Chronobiology International*, *25*, 597-609. doi:10.1080/07420520802261838

Chellappa, S., Morris, C., & Scheer, F. (2018). Daily circadian misalignment impairs human cognitive performance task-dependently. *Scientific Reports*, *8,* 1–11. doi:10.1038/s41598-018-20707-4

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics - Simulation and Computation*, *39*, 860-864. doi:10.1080/03610911003650383

Chun, M., & Turk-Browne, N. (2007). Interactions between attention and memory. *Current Opinion in Neurobiology*, *17*, 177-184. doi:10.1016/j.conb.2007.03.005

Clark, S., Benjamin, A., Wixted, J., Mickes, L., & Gronlund, S. (2015). Eyewitness identification and the accuracy of the criminal justice system. *Policy Insights from the Behavioral and Brain Sciences*, *2*, 175–186. doi:10.1177/2372732215602267

Cousins, J., Sasmita, K., & Chee, M. (2017). Memory encoding is impaired after multiple nights of partial sleep restriction. *Journal of Sleep Research*, *27*, 138-145. doi:10.1111/jsr.12578

Craik, F. I. M. (2002). Levels of processing: Past, present... and future? *Memory, 10*, 305–318. doi:10.1080/09658210244000135

Craik, F. I. M. (1999). Levels of encoding and retrieval. In: Challis, B. H., Velichkovskii, B. H. (Eds.). *Stratification in Cognition and Consciousness* (pp. 97–104). Amsterdam: John Benjamins. doi:10.1075/aicr.15.09cra

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behaviour, 11,* 671–684. doi:1016/S0022-5371(72)80001-X

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*, 268-294. doi:10.1037//0096-3445.104.3.268

Craik, F. I., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, *125*, 159-180. doi:10.1037/0096-3445.125.2.159

Crump, M., McDonnell, J., & Gureckis, T. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioural research. *Plos ONE*, 8, e57410. doi:10.1371/journal.pone.0057410

Czeisler, C., & Gooley, J. (2007). Sleep and Circadian Rhythms in Humans. *Cold Spring Harbor Symposia on Quantitative Biology*, *72*, 579-597. doi:10.1101/sqb.2007.72.064

Deffenbacher, K. (1980). Eyewitness accuracy and confidence: Can we infer anything about their relationship? *Law and Human Behavior*, *4*, 243–260. doi:10.1007/bf01040617

Deffenbacher, K., Bornstein, B., & Penrod, S. (2006). Mugshot exposure effects: Retroactive interference, mugshot commitment, source confusion, and unconscious transference. *Law and Human Behavior*, *30*, 287-307. doi:10.1007/s10979-006-9008-1

Di Milia, L., Adan, A., Natale, V., & Randler, C. (2013). Reviewing the Psychometric Properties of Contemporary Circadian Typology Measures. *Chronobiology International*, *30*, 1261-1271. doi:10.3109/07420528.2013.817415

Duffy, J. F., Rimmer, D. W., Czeisler, C. A. (2001). Association of intrinsic circadian period with morningness-eveningness, usual wake time, and circadian phase. *Behavioural Neuroscience, 115*, 895-899. doi:10.1037/0735-7044.115.4.895

Dunning, D., & Stern, L. (1994). Distinguishing accurate from inaccurate eyewitness identifications via inquiries about decision processes. *Journal of Personality and Social Psychology*, *67*, 818–835. doi:10.1037//0022-3514.67.5.818

Eckel-Mahan, K., & Storm, D. (2009). Circadian rhythms and memory: not so simple as cogs and gears. *EMBO Reports*, *10*, 584-591. doi:10.1038/embor.2009.123

Fabbri, M., Antonietti, A., Giorgetti, M., Tonetti, L., & Natale, V. (2007). Circadian typology and style of thinking differences. *Learning and Individual Differences*, *17*, 175-180. doi:10.1016/j.lindif.2007.05.002

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,* 1149-1160. doi: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*, 175-191. doi:10.3758/bf03193146

Felson, M., & Poulsen, E. (2003). Simple indicators of crime by time of day. *International Journal of Forecasting*, *19*, 595-601. doi:10.1016/s0169-2070(03)00093-1

Fernandes, M. A., & Moscovitch, M. (2000). Divided attention and memory: evidence of substantial interference effects at retrieval and encoding. *Journal of Experimental Psychology: General, 129,* 155-176. doi:10.1037//0096-3445.129.2.155

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*. doi:10.3389/fneur.2018.00056

Fitzgerald, R. J., Price, H. L., Oriet, C., & Charman, S. D. (2013). The effect of suspect-filler similarity on eyewitness identification decisions: A meta-analysis. *Psychology, Public Policy, and Law, 19*, 151–164. doi:10.1037/a0030618

Fitzgerald, R., & Price, H. (2015). Eyewitness identification across the life span: A meta-analysis of age differences. *Psychological Bulletin*, *141*, 1228-1265. doi:10.1037/bul0000013

Fleischer, A., Mead, A., & Huang, J. (2015). Inattentive responding in MTurk and other online samples. *Industrial and Organizational Psychology*, *8*, 196-202. doi:10.1017/iop.2015.25

Flowe, H., Colloff, M., Karoğlu, N., Zelek, K., Ryder, H., Humphries, J., & Takarangi, M. (2017). The effects of alcohol intoxication on accuracy and the confidence-accuracy relationship in photographic simultaneous line-ups. *Applied Cognitive Psychology*, *31*, 379–391. doi:10.1002/acp.3332

Forst, B. (2013). Wrongful convictions in a world of miscarriages of justice. In C.R. Huff and M. Killias (Eds.), *Wrongful Convictions and Miscarriages of Justice: Causes and remedies in North American and European criminal justice systems* (pp. 15–43). London: Routledge.

Gabbert, F., Hope, L., & Confrey, M. (2018). Witness testimony. In A. Griffiths, & R. Milne (Eds.), *The Psychology of Criminal Investigation: From Theory to Practice.* Routledge.

Gaggioni, G., Maquet, P., Schmidt, C., Dijk, D., & Vandewalle, G. (2014). Neuroimaging, cognition, light and circadian rhythms. *Frontiers in Systems Neuroscience*, *8*, 1–12. doi:10.3389/fnsys.2014.00126

Glass, L. (2001). Synchronization and rhythmic processes in physiology. *Nature*, *410*, 277-284. doi:10.1038/35065745

Goodsell, C. A., Neuschatz, J. S., & Gronlund, S. D. (2009). Effects of mugshot commitment on lineup performance in young and older adults. *Applied Cognitive Psychology, 23*, 788–803. doi:10.1002/acp.1512

Green, D., & Swets, J. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Halberg, F., Cornélissen, G., Katinas, G., Syutkina, E. V., Sothern, R. B., Zaslavskaya, R., ... & Tarquini, R. (2003). Transdisciplinary unifying implications of circadian findings in the 1950s. *Journal of circadian rhythms*, *1*, 2.

Hasher, L., & Zacks, R. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist*, *39*, 1372–1388. doi:10.1037/0003-066x.39.12.1372

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108*, 356–388. doi:10.1037/0096-3445.108.3.356

Hastie, R., & Kumar, P. A. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology, 37*, 25-38. doi:10.1037/0022-3514.37.1.25

Hastings, M., Reddy, A., & Maywood, E. (2003). A clockwork web: Circadian timing in brain and periphery, in health and disease. *Nature Reviews Neuroscience*, *4*, 649-661. doi:10.1038/nrn1177

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*, 2425-2430. doi:10.1126/science.1063736

Haxby, J., Hoffman, E., & Gobbini, M. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry*, *51*, 59-67. doi:10.1016/s0006-3223(01)01330-0

Hirst, W., & Kalmar, D. (1987). Characterizing attentional resources. *Journal of Experimental Psychology: General*, *116*, 68-81. doi:10.1037/0096-3445.116.1.68

Horne, J. A., & Östberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology, 4,* 97–100. doi:10.1177/0748730405285278

Ingram, K., Ay, A., Kwon, S., Woods, K., Escobar, S., & Gordon, M., … Jain, K. (2016). Molecular insights into chronotype and time-of-day effects on decision-making. *Scientific Reports*, *6,* 1–9. doi:10.1038/srep29392

Innocence Project. (2019). http://www.innocenceproject.org/Last visited September 29, 2019.

Intons-Peterson, M. J., Rocchi, P., West, T., McLellan, K., & Hackney, A. (1999). Age, testing at preferred or nonpreferred times (testing optimality), and false memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 23–40. doi:10.1037/0278-7393.25.1.23 10.2307/2531734

Jacoby, L. L. (1998). Invariance in automatic influences of memory: Toward a user's guide for the process-dissociation procedure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 3–26. doi:10.1037/0278-7393.24.1.3

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110*, 306–340. doi:10.1037/0096-3445.110.3.306

Jacoby, L. L., Woloshyn, V., & Kelley, C. M. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of experimental psychology: General, 118,* 115–125. doi:10.1037/0096-3445.118.2.115

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3–28. doi:10.1037/0033-2909.114.1.3

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 336–358. doi:10.1037//0278-7393.26.2.336

Kanwisher, N., McDermott, J., & Chun, M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience, 17*, 4302-4311. doi:10.1523/jneurosci.17-11-04302.1997

Kebbell, M. R., & Milne, R. (1998). Police officers’ perceptions of eyewitness performance in forensic investigations. *The Journal of Social Psychology*, 138, 323–330. doi:10.1080/00224549809600384

Kelley, P., Lockley, S., Foster, R., & Kelley, J. (2014). Synchronizing education to adolescent biology: ‘Let teens sleep, start school later’. *Learning, Media and Technology*, *40*, 210-226. doi:10.1080/17439884.2014.942666

Kihlstrom, J. (2009). ‘So that we might have roses in December’: The functions of autobiographical memory. *Applied Cognitive Psychology*, *23*, 1179-1192. doi:10.1002/acp.1618

Kihlstrom, J. F. (1996). Memory research: The convergence of theory and practice. In D. Hermann, M. Johnson, C. McEvoy, C. Hertzog, & P. Hertel (Eds.), *Basic and applied memory: Theory in context* (Vol. 1, pp. 5–25). Mahwah, NJ: Erlbaum

Krishnan, H., & Lyons, L. (2015). Synchrony and desynchrony in circadian clocks: Impacts on learning and memory. *Learning & Memory*, *22*, 426-437. doi:10.1101/lm.038877.115

Kyriacou, C., & Hastings, M. (2010). Circadian clocks: Genes, sleep, and cognition. *Trends in Cognitive Sciences*, *14*, 259–267. doi:10.1016/j.tics.2010.03.007

Levandovski, R., Sasso, E., & Hidalgo, M. (2013). Chronotype: A review of the advances, limits and applicability of the main instruments used in the literature to assess human phenotype. *Trends in Psychiatry and Psychotherapy*, *35*, 3–11. doi:10.1590/s2237-60892013000100002

Lockhart, R. S. (1992). Levels of processing. In: L. R. Squire (Ed.), *Encyclopedia of Learning and Memory.* New York, NY: Macmillan.

Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology*, *7*, 560-572. doi:10.1016/0010-0285(75)90023-7

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory, 12*, 361–366. doi:10.1101/lm.94705

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, *13*, 585-589. doi:10.1016/S0022-5371(74)80011-3

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 19–31. doi:10.1037/0278-7393.4.1.19

Loftus, E. F. (2003). Our changeable memories: Legal and practical implications. *Nature Reviews: Neuroscience, 4,* 231–234. doi:10.1038/nrn1054

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska Directed Emotional Faces - KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet*. ISBN 91-630-7164-9.

MacLin, O., MacLin, M., & Malpass, R. (2001). Race, arousal, attention, exposure and delay: An examination of factors moderating face recognition. *Psychology, Public Policy, and Law, 7*, 134–152. doi:10.1037/1076-8971.7.1.134

Madon, S., Guyll, M., Scherr, K., Greathouse, S., & Wells, G. (2012). Temporal discounting: The differential effect of proximal and distal consequences on confession decisions. *Law and Human Behavior*, *36*, 13–20. doi:10.1037/h0093962

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*, 252–271. doi:10.1037//0033-295x.87.3.252

Mason, W., & Suri, S. (2011). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*, 1-23. doi:10.3758/s13428-011-0124-6

May, C. (1999). Synchrony effects in cognition: The costs and a benefit. *Psychonomic Bulletin & Review*, *6*, 142–147. doi:10.3758/bf03210822

May, C. P., Hasher, L., & Foong, N. (2005). Implicit memory, age, and time of day: Paradoxical priming effects. *Psychological Science*, *16*, 96–100. doi: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*, 326–330. doi:10.1111/j.1467-9280.1993.tb00573.x

May, C., & Hasher, L. (1998). Synchrony effects in inhibitory control over thought and action. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 363–379. doi:10.1037//0096-1523.24.2.363

Meade, A., & Craig, S. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*, 437-455. doi:10.1037/a0028085

Meissner, C., & Brigham, J. (2001). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology*, *15*, 603–616. doi:10.1002/acp.728

Meissner, C., Tredoux, C., Parker, J., & MacLin, O. (2005). Eyewitness decisions in simultaneous and sequential lineups: A dual-process signal detection theory analysis. *Memory & Cognition*, *33*, 783–792. doi:10.3758/bf03193074

Memon, A., Hope, L., & Bull, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. *British Journal of Psychology*, *94*, 339-354. doi:10.1348/000712603767876262

Memon, A., Hope, L., Bartlett, J., & Bull, R. (2002). Eyewitness recognition errors: The effects of mugshot viewing and choosing in young and old adults. *Memory & Cognition*, *30*, 1219-1227. doi:10.3758/bf03213404

Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin, 135*, 638–677. doi:10.1037/a0015849

Monk, T. (1992). Circardian rhythms in subjective activation, mood, and performance efficiency. In M. Kryger, T. Roth, & W. Dement (Eds.). *Principles and practice of sleep medicine*, 2nd Ed. (pp. 321–330). Philadelphia: WB Saunders.

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*, 659–-667. doi:10.1017/S0033291706007252

Mulligan, N. W. (1998). The role of attention during encoding in implicit and explicit memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 27. doi:10.1037/0278-7393.24.1.27

Natale, V., & Cicogna, P. (2002). Morningness-eveningness dimension: Is it really a continuum? *Personality and Individual Differences, 32*, 809–816. doi:10.1016/S0191-8869(01)00085-X

Natale, V., Esposito, M., Martoni, M., & Fabbri, M. (2006). Validity of the reduced version of the Morningness-Eveningness Questionnaire. *Sleep and Biological Rhythms*, *4*, 72-74. doi:10.1111/j.1479-8425.2006.00192.x

Necka, E. (1997). Attention, working memory and arousal: Concepts apt to account for the "process of intelligence". In G. Matthews (Ed.), *Cognitive Science Perspectives on Personality and Emotion* (pp. 503-554). North-Holland: Elsevier. doi:10.1016/S0166-4115(97)80129-X

Nowack, K., & Van Der Meer, E. (2018). Nowack, K., Meer, E. Van Der, & Nowack, K. (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*, *36*, 1647-1662. doi:10.1080/07420528.2018.1500477

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*, 867-872.

Otgaar, H., Howe, M. L., Peters, M., Smeets, T., & Moritz, S. (2014). The production of spontaneous false memories across childhood. *Journal of Experimental Child Psychology, 121,* 28–41. doi.org:10.1016/j.jecp.2013.11.019

Otten, L. J., Henson, R. N., & Rugg, M. D. (2001). Depth of processing effects on neural correlates of memory encoding: Relationship between findings from across- and within-task comparisons. *Brain: A Journal of Neurology, 124,* 399–412. doi:10.1093/brain/124.2.399

Palermo, R., & Rhodes, G. (2002). The influence of divided attention on holistic face perception. *Cognition*, *82*, 225-257. doi:10.1016/S0010-0277(01)00160-3

Palmer, M. A., Brewer, N., Weber, N., & Nagesh, A. (2013). The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. *Journal of Experimental Psychology: Applied*, *19*, 55–71. doi:10.1037/a0031602

Paolacci, G., & Chandler, J. (2014). Inside the Turk. *Current Directions in Psychological Science*, *23*, 184-188. doi:10.1177/0963721414531598

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5*, 411-419.

Peek, C., Levine, D., Cedernaes, J., Taguchi, A., Kobayashi, Y., & Tsai, S. et al. (2017). Circadian clock interaction with HIF1α mediates oxygenic metabolism and anaerobic glycolysis in skeletal muscle. *Cell Metabolism*, *25*, 86-92. doi:10.1016/j.cmet.2016.09.010

Perfect, T.J., & Harris, L.J. (2003) Adult age differences in unconscious transference: Source confusion or identity blending?. *Memory & Cognition 31*, 570–580. doi:10.3758/BF03196098

Peters, M., Engel, M., Hauschildt, M., Moritz, S., Jelinek, L., & Otgaar, H. (2012). Investigating the corrective effect of forewarning on memory and meta-memory deficits in schizophrenia patients. *Journal of Experimental Psychopathology*, *3*, 673-687. doi:10.5127/jep.022011

Peterson, J., Sommers, I., Baskin, D., & Johnson, D. (2010). *The role and impact offorensic evidence in the criminal justice process.* Revised final report. Washington, DC: National Institute of Justice.

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*, 63–72. doi:10.1111/j.2044-8295.1990.tb02346.x

Puttaert, D., Adam, S., & Peigneux, P. (2018). Subjectively-defined optimal / non-optimal time of day modulates controlled but not automatic retrieval processes in verbal memory. *Journal of Sleep Research, 28,* e12798. doi:10.1111/jsr.12798

Rand, D. (2012). The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology, 299*, 172-179. doi:10.1016/j.jtbi.2011.03.004

Rasch, B., & Born, J. (2013). About sleep's role in memory. *Physiological Reviews*, *93*, 681-766. doi:10.1152/physrev.00032.2012

Read, J. D., Tollestrup, P., Hammersley, R., McFadzen, E., & Christensen, A. (1990). The unconscious transference effect: Are innocent bystanders ever misidentified? *Applied Cognitive Psychology, 4*, 3–31. doi:10.1002/acp.2350040103

Refinetti, R. (2006). Circadian physiology (2nd ed.). CRC Press.

Reppert, S. M., & Weaver, D. R. (2002). Coordination of circadian timing in mammals. *Nature*, *418*, 935-941. doi:10.1038/nature00965

Richler, J. J., & Gauthier, I. (2014). A meta-analysis and review of holistic face processing. *Psychological Bulletin*, *140*, 1281–1302. doi:10.1037/a0037004

Ridley, A. M., Gabbert, F., & La Rooy, D. J. (Eds.). (2013). *Wiley series in the psychology of crime, policing and law. Suggestibility in legal contexts: Psychological research and forensic implications.* Wiley-Blackwell.

Riley, E., Esterman, M., Fortenbaugh, F., & DeGutis, J. (2017). Time-of-day variation in sustained attentional control. *Chronobiology International*, *34*, 993–1001. doi:10.1080/07420528.2017.1308951

Robotham, R. J., & Starrfelt, R. (2017). Face and word recognition can be selectively affected by brain injury or developmental disorders. *Frontiers in Psychology*, *8*, 1547. doi:10.3389/fpsyg.2017.01547

Roediger, H., & McDermott, K. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 803–814. doi:10.1037//0278-7393.21.4.803

Roediger, H., Watson, J., McDermott, K., & Gallo, D. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review*, *8*, 385–407. doi:10.3758/bf03196177

Roenneberg, T., Kuehnle, T., Juda, M., Kantermann, T., Allebrandt, K., Gordijn, M., & Merrow, M. (2007). Epidemiology of the human circadian clock. *Sleep Medicine Reviews, 11*, 429–438. doi:10.1016/j.smrv.2007.07.005

Roenneberg, T., Wirz-Justice, A., & Merrow, M. (2003). Life between clocks: Daily temporal patterns of human chronotypes. *Journal of Biological Rhythms, 18*, 80–90. doi:10.1177/0748730402239679

Ross, D. R., Ceci, S. J., Dunning, D., & Toglia, M. P. (1994). Unconscious transference and mistaken identity: When a witness misidentifies a familiar but innocent person. *Journal of Applied Psychology, 79*, 918–930. doi:10.1037/0021-9010.79.6.918

Ryan, L., Hatfield, C., & Hofstetter, M. (2002). Caffeine reduces time-of-day effects on memory performance in older adults. *Psychological Science*, *13*, 68–71. doi:10.1111/1467-9280.00412

Sauer, J., & Hope, L. (2016). The effects of divided attention at study and reporting procedure on regulation and monitoring for episodic recall. *Acta Psychologica*, *169*, 143-156. doi:10.1016/j.actpsy.2016.05.015

Sauerland, M., & Sporer, S. (2007). Post-decision confidence, decision time, and self-reported decision processes as postdictors of identification accuracy. *Psychology, Crime & Law*, *13*, 611–625. doi:10.1080/10683160701264561

Sauerland, M., & Sporer, S. L. (2009). Fast and confident: Postdicting eyewitness identification accuracy in a field study. *Journal of Experimental Psychology: Applied, 15,* 46–62. doi:10.1037/a0014560

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,* 978–992. doi:10.3758/s13421-014-0401-6

Schacter, D. L., Kaszniak, A. W., Kihlstrom, J. F., & Valdiserri, M. (1991). The relation between source memory and aging. *Psychology and Aging, 6*, 559–568. doi:10.1037/0882-7974.6.4.559

Scherr, K., Miller, J., & Kassin, S. (2014). “Midnight confessions”: The effect of chronotype asynchrony on admissions of wrongdoing. *Basic and Applied Social Psychology*, *36*, 321–328. doi:10.1080/01973533.2014.917974

Schmidt, C., Collette, F., Cajochen, C., & Peigneux, P. (2007). A time to think: Circadian rhythms in human cognition. *Cognitive Neuropsychology*, *24*, 755–789. doi:10.1080/02643290701754158

Schoenfield, D., and Stones, M. J. (1979). Remembering and aging. In Kihlstrom, J. F., and Evans, F. J. (Eds.), *Functional Disorders of Memory.* Erlbaum, Hillsdale, New Jersey

Schonfield, D., & Robertson, B. A. (1966). Memory storage and aging. *Canadian Journal of Psychology/Revue canadienne de psychologie, 20*, 228–236. doi:10.1037/h0082941

Schwartz, B. L. (2014). Memory for people: Integration of face, voice, name, and biographical information. In T. J. Perfect & D. S. Lindsay (Eds.), *The SAGE Handbook of Applied Memory* (pp. 3–19). London, UK: SAGE. doi:10.4135/9781446294703.n1

Sim, J., & Wright, C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy*, *85*, 257-268. doi:10.1093/ptj/85.3.257

Smith, C., Reilly, C., & Midkiff, K. (1989). Evaluation of three circadian rhythm questionnaires with suggestions for an improved measure of morningness. *Journal of Applied Psychology*, *74*, 728-738. doi:10.1037//0021-9010.74.5.728

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4*, 108–131. doi:10.1207/S15327957PSPR0402_01

Smolensky, M. H., & Peppas, N. A. (2007). Chronobiology, drug delivery, and chronotherapeutics. *Advanced Drug Delivery Reviews*, *59*, 828-851.

Snider, K., Sullivan, K., & Obrietan, K. (2018). Circadian Regulation of Hippocampal-Dependent Memory: Circuits, Synapses, and Molecular Mechanisms. *Neural Plasticity*, *2018*, 1-13. doi:10.1155/2018/7292540

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*, 34–50. doi:10.1037/0096-3445.117.1.34

Sporer, S., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin*, *118*, 315–327. doi:10.1037/0033-2909.118.3.315

Steblay, N. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law and Human Behavior*, *21*, 283–297. doi:10.1023/a:1024890732059

St-Yves, M. (2015). Investigative Interviewing: *The Essential Handbook of Best Practices.* Canada: Carswell.

Tanaka, J., & Farah, M. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology Section A*, *46*, 225-245. doi:10.1080/14640749308401045

Thayer, R. (1987). Problem perception, optimism, and related states as a function of time of day (diurnal rhythm) and moderate exercise: Two arousal systems in interaction. *Motivation and Emotion*, *11*, 19-36. doi:10.1007/bf00992211

Torsvall, L., & Åkerstedt, T. (1980). A diurnal type scale. Construction, consistency and validation in shift work. *Scandinavian Journal of Work, Environment & Health*, *6*, 283-290. doi:10.5271/sjweh.2608

Tredoux, C. (1999). Statistical considerations when determining measures of lineup size and lineup bias. *Applied Cognitive Psychology, 13,* 9–26. doi:10.1002/(SICI)1099-0720

Trott, C. T., Friedman, D., Ritter, W., Fabiani, M., & Snodgrass, J. G. (1999). Episodic priming and memory for temporal source: Event-related potentials reveal age-related differences in prefrontal functioning. *Psychology and Aging, 14*, 390–413. doi:10.1037/0882-7974.14.3.390

Tulving, E. (1962). Subjective organization in free recall of" unrelated" words. *Psychological review, 69,* 344-354. doi:10.1037/h0043150

Valdez P. (2019). Circadian Rhythms in Attention. *The Yale Journal of Biology and Medicine*, *92*, 81–92.

Valdez, P. (2019). Homeostatic and circadian regulation of cognitive performance. *Biological Rhythm Research, 50*, 85-93. doi:10.1080/09291016.2018.1491271

Van Dongen H., & Dinges D. F. (2003). Investigating the interaction between the homeostatic and circadian processes of sleep–wake regulation for the prediction of waking neurobehavioural performance. *Journal of Sleep Research, 12,* 181–187. doi:10.1046/j.1365-2869.2003.00357.x

van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences, 35*, 211–219. doi:10.1016/j.tins.2012.02.001

Wagenaar, W. A. (2009). Expert witness in international war crimes tribunals. *Psychology, Crime & Law*, *15*, 583-596. doi:10.1080/10683160802438338

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*, 321–332. doi:10.1080/10683169608409787

Weber, N., & Brewer, N. (2003). The effect of judgment type and confidence scale on confidence-accuracy calibration in face recognition. *Journal of Applied Psychology*, 88, 490–499.

Weber, N., & Brewer, N. (2004). Confidence-accuracy calibration in absolute and relative face recognition judgments. *Journal of Experimental Psychology: Applied*, *10*, 156–172. doi:10.1037/1076-898x.10.3.156

Weber, N., & Brewer, N. (2006). Positive versus negative face recognition decisions: Confidence, accuracy, and response latency. *Applied Cognitive Psychology, 20,* 17–31. doi:10.1002/acp.1166

Weber, N., Brewer, N., Wells, G. L., Semmler, C., & Keast, A. (2004). Eyewitness identification accuracy and response latency: The unruly 10–12 second rule. *Journal of Experimental Psychology: Applied, 10,* 139–147. doi:10.1037/1076-898X.10.3.139

Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology, 36*, 1546–1557. doi:10.1037/0022-3514.36.12.1546

Wells, G. L., & Penrod, S. D. (2011). Eyewitness identification research: Strengths and weaknesses of alternative methods. In B. Rosenfeld & S. D. Penrod

(Eds.), *Research methods in forensic psychology* (pp. 237-256). Hoboken, NJ, US: John Wiley & Sons Inc.

Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest*, *7*, 45-75. doi:10.1111/j.1529-1006.2006.00027.x

Wells, G., & Olson, E. (2003). Eyewitness Testimony. *Annual Review of Psychology*, *54*, 277-295. doi:10.1146/annurev.psych.54.101601.145028

Wells, G.L. (2014). Eyewitness identification: Probative value, criterion Shifts, and policy regarding the sequential lineup. *Current Directions in Psychological Science, 23,* 11–16. doi:10.1177/0963721413504781

Wessling, K. S., Huber, J., & Netzer, O. (2017). MTurk character misrepresentation: Assessment and solutions. *Journal of Consumer Research*, *44*, 211–230. doi:10.1093/jcr/ucx053

Wilson J.P., Bernstein M.J., & Hugenberg K. (2016) A synthetic perspective on the own-race bias in eyewitness identification. In Bornstein B., & Miller M. (Eds), *Advances in psychology and law, Vol. 2* (pp 241-270). doi:10.1007/978-3-319-43083-6_8

Wixted, J., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review*, *121*, 262-276. doi:10.1037/a0035940

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*, 19–43. doi: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*, 508–521. doi:10.1162/jocn.2006.18.4.508

Wright, K., Lowry, C., & LeBourgeois, M. (2012). Circadian and wakefulness-sleep modulation of cognition in humans. *Frontiers in Molecular Neuroscience*, *5,* 1–12. doi:10.3389/fnmol.2012.00050

Yang, L., Hasher, L., & Wilson, D. (2007). Synchrony effects in automatic and controlled retrieval. *Psychonomic Bulletin & Review*, *14*, 51–56. doi:10.3758/bf03194027

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*, 441–517. doi:10.1006/jmla.2002.2864

Yonelinas, A. P., & Jacoby, L. L. (2012). The process-dissociation approach two decades later: Convergence, boundary conditions, and new directions. *Memory & Cognition, 40*, 663–680. doi:10.3758/s13421-012-0205-5

Yoon, C. (1997). Age differences in consumers' processing strategies: An investigation of moderating influences. *Journal of Consumer Research*, *24*, 329–342. doi:10.1086/209514

Zeger, S., Liang, K., & Albert, P. (1988). Models for longitudinal data: A Generalized Estimating Equation approach. *Biometrics*, *44*, 1049–1060. 10.2307/2531734

Appendices

List of Appendices

Appendix A: Cued questions about the event and people involved (Chapter 2; Experiment 1)

Appendix A

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 color.

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?

Appendix B

Supplementary Table 1

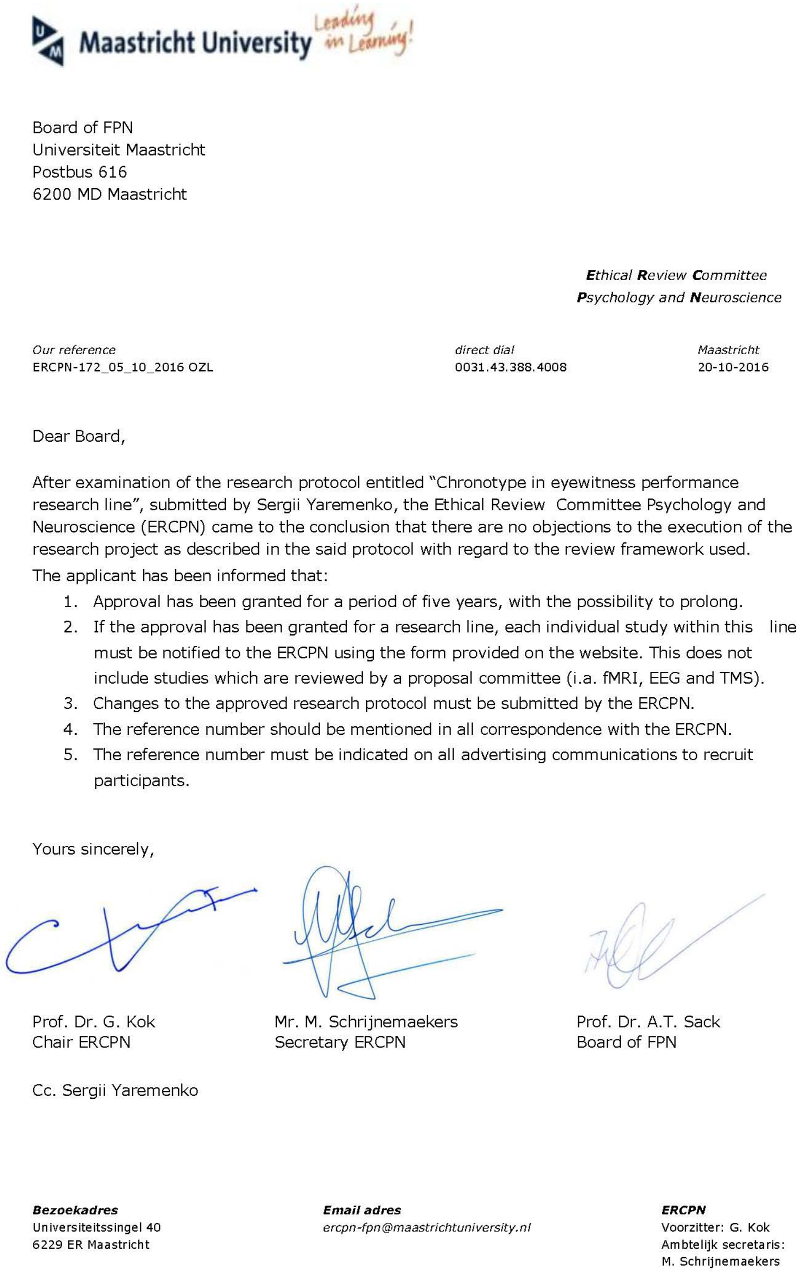
Counterbalancing face sets in Experiment 3

Optimal								**Non-optimal**							
Criminal scenario – Set 1 Supermarket scenario – Set 2				Criminal scenario – Set 2 Supermarket scenario – Set 1				Criminal scenario – Set 1 Supermarket scenario – Set 2				Criminal scenario – Set 2 Supermarket scenario – Set 1			
Task 1	***Task 1***	***Task 1***	***Task 1***	***Task 1***	***Task 1***	***Task 1***	***Task 1***	***Task 1***	***Task 1***	***Task 1***	***Task 1***	***Task 1***	***Task 1***	***Task 1***	***Task 1***
S1SsA	S1SsB	S1SsA	S1SsB	S2SsA	S2SsB	S2SsA	S2SsB	S1SsA	S1SsB	S1SsA	S1SsB	S2SsA	S2SsB	S2SsA	S2SsB
S2SsA	S2SsB	S2SsB	S2SsA	S1SsA	S1SsB	S1SsB	S1SsA	S2SsA	S2SsB	S2SsB	S2SsA	S1SsA	S1SsB	S1SsB	S1SsA
New1	New2	New1	New2	New1	New2	New1	New2	New1	New2	New1	New2	New1	New2	New1	New2
Task 2	***Task 2***	***Task 2***	***Task 2***	***Task 2***	***Task 2***	***Task 2***	***Task 2***	***Task 2***	***Task 2***	***Task 2***	***Task 2***	***Task 2***	***Task 2***	***Task 2***	***Task 2***
S1SsB	S1SsA	S1SsB	S1SsA	S2SsB	S2SsA	S2SsB	S2SsA	S1SsB	S1SsA	S1SsB	S1SsA	S2SsB	S2SsA	S2SsB	S2SsA
S2SsB	S2SsA	S2SsA	S2SsB	S1SsB	S1SsA	S1SsA	S1SsB	S2SsB	S2SsA	S2SsA	S2SsB	S1SsB	S1SsA	S1SsA	S1SsB
New2	New1	New2	New1	New2	New1	New2	New1	New2	New1	New2	New1	New2	New1	New2	New1
Group 1	**Group 2**	**Group 3**	**Group 4**	**Group 5**	**Group 6**	**Group 7**	**Group 8**	**Group 9**	**Group 10**	**Group 11**	**Group 12**	**Group 13**	**Group 14**	**Group 15**	**Group 16**

Note. Task 1 – face recognition task, Task 2 – source memory task, S1SsA – Set 1 Subset A, S2SsA – Set 2 Subset A, S1SsB – Set 1 Subset B, S2SsB –Set 2 Subset B, New1 – distractor faces (Set 1), New2 – distractor faces (Set 2).

Appendix C - Ethical Approval

The protocol of experiments presented in this thesis was examined and approved by the Ethical Review Committee Psychology and Neuroscience (ERCPN) of Maastricht University. Below is the official letter of approval of the research line. Each individual study within the research line received individual approval from Maastricht University ECP.



Maastricht University Leading in Learning!

Board of FPN
Universiteit Maastricht
Postbus 616
6200 MD Maastricht

Ethical ***R***eview ***C***ommittee
Psychology and ***N***euroscience

Our reference	*direct dial*	*Maastricht*
ERCPN-172_05_10_2016 OZL	0031.43.388.4008	20-10-2016

Dear Board,

After examination of the research protocol entitled "Chronotype in eyewitness performance research line", submitted by Sergii Yaremenko, the Ethical Review Committee Psychology and Neuroscience (ERCPN) came to the conclusion that there are no objections to the execution of the research project as described in the said protocol with regard to the review framework used.

The applicant has been informed that:

1. Approval has been granted for a period of five years, with the possibility to prolong.
2. If the approval has been granted for a research line, each individual study within this line must be notified to the ERCPN using the form provided on the website. This does not include studies which are reviewed by a proposal committee (i.a. fMRI, EEG and TMS).
3. Changes to the approved research protocol must be submitted by the ERCPN.
4. The reference number should be mentioned in all correspondence with the ERCPN.
5. The reference number must be indicated on all advertising communications to recruit participants.

Yours sincerely,

Prof. Dr. G. Kok	Mr. M. Schrijnemaekers	Prof. Dr. A.T. Sack
Chair ERCPN	Secretary ERCPN	Board of FPN

Cc. Sergii Yaremenko

Bezoekadres
Universiteitssingel 40
6229 ER Maastricht

Email adres
ercpn-fpn@maastrichtuniversity.nl

ERCPN
Voorzitter: G. Kok
Ambtelijk secretaris:
M. Schrijnemaekers

Appendix D – UPR16 Form

FORM UPR16

Research Ethics Review Checklist

Please include this completed form as an appendix to your thesis (see the Research Degrees Operational Handbook for more information

Postgraduate Research Student (PGRS) Information		**Student ID:**	838007
PGRS Name:	Sergii Yaremenko		
Department:	Psychology	**First Supervisor:**	Prof. Lorraine Hope
Start Date: (or progression date for Prof Doc students)		**28 August 2016**	

Study Mode and Route:	Part-time ☐ Full-time ☒	MPhil ☐ PhD ☒	MD ☐ Professional Doctorate ☐

Title of Thesis:	Time-of-Day Optimality Effects on Eyewitness Memory Performance
Thesis Word Count: (excluding ancillary data)	

If you are unsure about any of the following, please contact the local representative on your Faculty Ethics Committee for advice. Please note that it is your responsibility to follow the University's Ethics Policy and any relevant University, academic or professional guidelines in the conduct of your study

Although the Ethics Committee may have given your study a favourable opinion, the final responsibility for the ethical conduct of this work lies with the researcher(s).

UKRIO Finished Research Checklist:
(If you would like to know more about the checklist, please see your Faculty or Departmental Ethics Committee rep or see the online version of the full checklist at: http://www.ukrio.org/what-we-do/code-of-practice-for-research/)

a) Have all of your research and findings been reported accurately, honestly and within a reasonable time frame?	YES ☒ NO ☐
b) Have all contributions to knowledge been acknowledged?	YES ☒ NO ☐
c) Have you complied with all agreements relating to intellectual property, publication and authorship?	YES ☒ NO ☐
d) Has your research data been retained in a secure and accessible form and will it remain so for the required duration?	YES ☒ NO ☐
e) Does your research comply with all legal, ethical, and contractual requirements?	YES ☒ NO ☐

Candidate Statement:

I have considered the ethical dimensions of the above named research project, and have successfully obtained the necessary ethical approval(s)

Ethical review number(s) from Faculty Ethics Committee (or from NRES/SCREC):	172_05_10_2016

If you have *not* submitted your work for ethical review, and/or you have answered 'No' to one or more of questions a) to e), please explain below why this is so:

Signed *(PGRS)*:		**Date:** 28 August 2019