

Cognitive Psychology

The Interplay of Time-of-day and Chronotype Results in No General and Robust Cognitive Boost

Alodie Rey-Mermet¹ unverified^a, Nicolas Rothen¹ unverified

¹ Faculty of Psychology, UniDistance Suisse, Brig, Switzerland

Keywords: circadian arousal, attentional control, executive functions, working memory, individual differences

<https://doi.org/10.1525/collabra.88337>

Collabra: Psychology

Vol. 9, Issue 1, 2023

Using physiologically validated questionnaires in which the peak of circadian arousal is determined through morningness-eveningness preferences, individuals can be categorized into morning or evening chronotypes. Typically, individuals with such chronotypes are assumed to show better cognitive performance at their subjective peak of circadian arousal than at off peak. Although this so-called synchrony effect is accepted as common knowledge, empirical evidence is rather mixed. This may be explained by two methodical challenges. First, most studies are underpowered. Second, they include one task, but tasks differ across studies. Here, we tested the synchrony effect by focusing on two cognitive constructs that are assumed to underlie a wide variety of behaviors, that is: short-term maintenance and attentional control. Short-term maintenance refers to our ability to maintain information temporarily. Attentional control refers to our ability to avoid being distracted by irrelevant information. We addressed the methodical challenges by asking 446 young adults to perform eight tasks at on- and off-peak times. Four tasks were used to assess temporary maintenance of information (i.e., short-term memory). Four tasks were used to assess temporary maintenance and manipulation of information (i.e., working memory). Using structural equation modeling, we modeled attentional control as the goal-directed nature of the working-memory tasks without their maintenance aspects. At the individual-task level, there was some evidence for a synchrony effect. However, the evidence was weak and limited to two tasks. Moreover, at the latent-variable level, the results showed no evidence for a robust and general synchrony effect. These results were observed for the full sample ($N = 446$) and the subsample including participants with moderate to definite morning or evening chronotypes ($N = 191$). We conclude that the synchrony effect is most likely a methodical artefact and discuss the impact of our research on psychological science and scientific research more widely.

Environmental conditions at day and night are so different that it is hard to imagine that they do not influence human behaviour and experience. Indeed, previous research has shown that morningness-eveningness preferences, which are established by questionnaires, can be used to classify individuals into different chronotypes (Griefahn et al., 2001; Horne & Östberg, 1976). Moreover, these chronotypes have been related to physiological measures of circadian arousal (e.g., Horne & Östberg, 1977). Thus, most individuals who report a morningness preference and thus are classified as morning types have their peak of circadian arousal in the morning. By contrast, most individuals who report an eveningness preference and thus are classified as evening types have their peak of circadian arousal in the evening. Common knowledge as well as a large body of lit-

erature in psychology and neurosciences suggest an interplay between chronotype and time-of-day (e.g., May et al., 1993; May, 1999; May et al., 2005; May & Hasher, 1998a). That is, individuals with morning and evening types are assumed to exhibit better cognitive performance at their peak than at off peak. However, due to methodical challenges, the evidence for this so-called synchrony effect is not as clear as it seems. Here, we addressed these methodical challenges to clarify whether a general and robust synchrony effect can be observed.

Although the synchrony effect is accepted as true and is not really questioned in daily life (see, e.g., *Brain Function of Night Owls and Larks Differ, Study Suggests*, 2019; Ceurstemont, 2020; Cohen, 2014; Pink, 2018; Savage, 2020), empirical evidence is more equivocal. This is partic-

^a Correspondence: alodie.rey-mermet@fernuni.ch

ularly the case for cognitive constructs for which a substantial synchrony effect has been put forward in early studies (see, e.g., Intons-Peterson et al., 1998; May, 1999; May & Hasher, 1998b; West et al., 2002). For instance, evidence is mixed for tasks measuring attentional control or executive functions (i.e., the ability to maintain goal-relevant information when facing distraction; Draheim et al., 2022; von Bastian et al., 2020). While some studies present evidence for the synchrony effect on attentional control (e.g., Bennett et al., 2008; Hahn et al., 2012; Hasher et al., 2002; Intons-Peterson et al., 1998; Lara et al., 2014; May, 1999; May & Hasher, 1998b), others are not able to detect such an effect on attentional control (e.g., Bennett et al., 2008; Heimola et al., 2021; Knight & Mather, 2013; Matchock & Mordkoff, 2008; May & Hasher, 1998b; Schmidt et al., 2012). The same situation applies to working memory (i.e., the ability to manipulate and maintain information for a short duration; e.g., Baddeley, 2012). That is, some studies find evidence for a synchrony effect on working memory (e.g., Rowe et al., 2009; Schmidt et al., 2015; West et al., 2002), while other studies suggest no synchrony effect on working memory (e.g., Ceglarek et al., 2021; Heimola et al., 2021; Lewandowska et al., 2017).

Based on this mixed evidence, it is tempting to conclude that the synchrony effect is not general and robust. However, the current evidence may be distorted by at least two methodical challenges. First, most studies investigating the synchrony effect on human cognition were statistically underpowered. Table 1 presents an overview of previous research about the sample size and the type of testing-time manipulation (i.e., whether on- and off-peak times were manipulated between-subjects or within-subject). We estimated the adequacy of the sample sizes by comparing them to the recommended sample sizes put forward by Brysbaert (2019). These recommended sample sizes were determined using a generic effect size of Cohen's d of .40. According to Brysbaert (2019), effect sizes from previous studies should be avoided to determine the target sample sizes because published studies are frequently underpowered and the impact of the publication bias is unknown. He suggests using a Cohen's d of .40, because this represents "a good first estimate of the smallest effect size of interest in psychological research" (see p. 1). Thus, with such a Cohen's d , an alpha level of .05, and a power of .80, Brysbaert (2019) showed that the sample size should consist of 200 participants for a t -test comparing on- and off-peak times in a between-subjects design and of 52 participants for a t -test comparing on- and off-peak times in a within-subject design. If a stricter criterium is applied for the power by using a cut-off .90, the sample-size requirements increase to 264 and 70 participants, respectively. According to these recommendations, 81% and 89%, respectively, of the studies listed in Table 1 included small sample sizes and thus had not enough statistical power to detect a true effect.

The second methodological challenge is that most studies included one task, and the tasks differed across the studies. This challenge concerns, for example, about 81% of the studies listed in Table 1. Only including one task per study is questionable because of the task-impurity prob-

lem. That is, each cognitive task measures not only the construct of interest but also other constructs and random noise (e.g., Miyake & Friedman, 2012). Thus, the mixed evidence regarding the synchrony effect may result from different studies controlling more or less well for this task-impurity problem. Moreover, including different tasks across different studies is problematic because these different tasks may not assess the same construct. For example, working memory is often synonymously used for short-term memory (i.e., the ability only to maintain information for a short duration; see Ceglarek et al., 2021; Lewandowska et al., 2017 for examples of such a mismatch). Thus, the mixed evidence regarding the synchrony effect in working memory may result from different constructs being measured. For attentional control, this issue is even greater. Recent research has emphasized that the different tasks used to assess attentional control are not measuring the same construct but rather task-specific processes (e.g., Karr et al., 2018; Rey-Mermet et al., 2018, 2019, 2020, 2021). Accordingly, the mixed evidence regarding the synchrony effect in attentional control may be the result of some task-specific processes being affected by the synchrony effect and other task-specific processes being affected by no synchrony effect. This makes the presence or the absence of the synchrony effect difficult to predict at a theoretical level, thus challenging the generality and robustness of this effect.

The goal of the present study was to determine the scope and robustness of the synchrony effect by focusing on two cognitive constructs that are assumed to underlie a wide variety of behaviors and experiences and by addressing the outlined methodical challenges. Specifically, we aimed to investigate the synchrony effect on working memory and attentional control using two sets of four tasks. In one set, the tasks required temporary maintenance of information (i.e., short-term memory). In the other set, the tasks required temporary maintenance and manipulation of information (i.e., working memory). Using structural equation modeling, we modeled attentional control as the goal-directed nature of the working-memory tasks without their maintenance aspects. The modeling approach also enables us to solve the task-impurity problem by capturing the constructs of interest (i.e., attentional control and short-term maintenance) as the shared variance across the measures. Furthermore, we used a within-subject design in which a large sample of participants were tested at both peak and off-peak times.

We hypothesized that if the impact of circadian arousal can be assessed as a synchrony effect on working memory and attentional control (e.g., May & Hasher, 1998b; Rowe et al., 2009; West et al., 2002), we should be able to find better performance at peak than at off-peak times for the short-term memory and working-memory tasks as well as for the latent constructs of short-term maintenance and attentional control. In contrast, if the impact of circadian arousal cannot be measured as a synchrony effect on working memory and attentional control (e.g., Matchock & Mordkoff, 2008; May & Hasher, 1998b), performance should not differ between peak and off-peak times at the individual-task level and the latent-variable level.

Table 1. Overview of the Sample Sizes and Design Used in Previous Research.

Study	Sample size	Number of task(s)	Testing time
<i>Bennett et al. (2008)</i>	77	> 1	<i>between-subjects</i>
<i>Bodenhausen (1990, Exp. 1)</i>	55	1	<i>between-subjects</i>
<i>Bodenhausen (1990, Exp. 2)</i>	189	1	<i>between-subjects</i>
<i>Ceglarek et al. (2021)</i>	66	1	<i>within-subject</i>
<i>Fabbri et al. (2013, Exp. 1)</i>	170	1	<i>between-subjects</i>
<i>Fabbri et al. (2013, Exp. 2)</i>	234	1	<i>between-subjects</i>
<i>Goldstein et al. (2007)</i>	80	> 1	<i>between-subjects</i>
<i>Hahn et al. (2012)</i>	80	> 1	<i>between-subjects</i>
<i>Hasher et al. (2002)</i>	96	1	<i>between-subjects</i>
<i>Intons-Peterson et al. (1998)</i>	64 / 40	1	<i>between-subjects</i>
<i>Intons-Peterson et al. (1999, Exp. 1)</i>	77 / 42	> 1	<i>between-subjects</i>
<i>Intons-Peterson et al. (1999, Exp. 3)</i>	90 / 67	> 1	<i>between-subjects</i>
<i>Lara et al. (2014)</i>	27	> 1	<i>within-subject</i>
<i>Lehmann et al. (2013)</i>	42 / 42	1	<i>between-subjects</i>
<i>Lewandowska et al. (2018)</i>	52	> 1	<i>within-subject</i>
<i>Li et al. (1998, Exp. 1)</i>	32 / 32	1	<i>between-subjects</i>
<i>Li et al. (1998, Exp. 2)</i>	32 / 31	1	<i>between-subjects</i>
<i>Matchock and Mordkoff (2009)</i>	80	1	<i>within-subject</i>
<i>May (1999)</i>	40 / 44	1	<i>between-subjects</i>
<i>May and Hasher (1998, Exp. 1)</i>	48 / 48	1	<i>between-subjects</i>
<i>May and Hasher (1998, Exp. 2)</i>	36 / 36	1	<i>between-subjects</i>
<i>May et al. (1993)</i>	20 / 18	1	<i>between-subjects</i>
<i>May et al. (2005, Exp. 1)</i>	36 / 48	1	<i>between-subjects</i>
<i>May et al. (2005, Exp. 2)</i>	54 / 36	1	<i>between-subjects</i>
<i>Petros et al. (1990)</i>	79	1	<i>between-subjects</i>
<i>Rothen and Meier (2016)</i>	160	1	<i>within-subject</i>
<i>Rothen and Meier (2017)</i>	115 / 113	1	<i>within-subject</i>
<i>Rowe et al. (2009)</i>	56 / 55	1	<i>between-subjects</i>
<i>Schmidt et al. (2012)</i>	31	1	<i>within-subject</i>
<i>Schmidt et al. (2015)</i>	28	1	<i>within-subject</i>
<i>Van Opstaal (2021)</i>	130	1	<i>within-subject</i>
<i>West et al. (2002)</i>	20 / 20	1	<i>within-subject</i>
<i>Yang et al. (2007, Exp. 1)</i>	0 / 52	1	<i>between-subjects</i>
<i>Yang et al. (2007, Exp. 2)</i>	0 / 46	1	<i>between-subjects</i>
<i>Yaremenko et al. (2021)</i>	91	1	<i>between-subjects</i>
<i>Yoon (1997)</i>	80 / 85	1	<i>between-subjects</i>

Note. For studies including young and/or older adults, both sample sizes are given separately. The first value refers to the sample size for young adults; the second value refers to the sample size for older adults. In all studies, the samples included participants that were categorized as morning and evening chronotypes. Only in Van Opstaal et al. (2022), the sample included participants with neutral chronotypes in addition to participants with morning and evening chronotypes. In the column "The number of task(s)", a study is considered to include more than one task if the tasks are assumed to measure the same construct. In the column "Testing time", a between-subjects manipulation of the testing time refers to a design in which one group of participants were tested at their subjective peak of circadian arousal and another group of participants were tested at their subjective off peak. A within-subject manipulation of the testing time refers to a design in which all participants were tested at their subjective on and off peaks. For the sake of clarity, studies which are considered as underpowered irrespective of whether the cut-off for the power was set to .80 or .90 are presented in bold and italic. Studies which are considered as underpowered when the cut-off was set to .90 only are presented in italic.

Methods

Transparency and openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study (Simmons et al., 2012). This study's design and its analysis were pre-registered (see <https://osf.io/tywu7>). All devia-

tions from the preregistration are presented in Appendix A. For all analyses, we used R (Version 4.3.1; R Core Team, 2021) and the R-packages *afex* (Version 1.3.0; Singmann et al., 2021), *BayesFactor* (Version 0.9.12.4.4; Morey, 2008), *ggplot2* (Version 3.4.2; Wickham, 2016), *lavaan* (Version 0.6.16; Rosseel, 2012), *papaja* (Version 0.1.1; Aust & Barth, 2020), *psych* (Version 2.3.6; Revelle, 2021), *semTools* (Ver-

sion 0.5.6; Jorgensen et al., 2021), and *splithalf* (Version 0.8.2; Parsons, 2020).

Participants

All participants were recruited and tested by the students from our university who took part in the course M08 in fall term 2020 and in the course M1 in spring term 2021. Participants were thus students' acquaintance. Before the recruitment, the students were informed about the most influential studies investigating the synchrony effect in attentional control and working memory (with, e.g., May & Hasher, 1998b; Rowe et al., 2009; West et al., 2002). To complete the course, the students were asked to test participants following best research practices, for example, by instructing the participants carefully and by protocolling the issues if some occurred. Therefore, for a student, successfully recruiting and testing participants was not tied to whether or not their participants showed a synchrony effect. Furthermore, it was tied neither to their grades nor to the exclusion criteria used in the present study. This was implemented to ensure no bias in recruitment and testing from the side of the students.

From the side of the participants, we avoided any selection bias by testing all participants who were willing to participate. Due to this data-collection procedure, the analyses were planned to be performed on all chronotypes (morning, evening, and neutral). Therefore, in the preregistration, we determined the target sample size for a sample including all chronotypes. Following Brysbaert's (2019) recommendations, we opted for a generic effect size of .196 – that is, a Cohen's *d* of .40 –, a statistical power of .90, and a probability level of .05. Furthermore, we hypothesized a measurement model including four latent variables and 16 manifest variables. Using the a-priori Sample Size Calculator for Structural Equation Models (retrieved from <http://www.danielsoper.com/statcalc>; Soper, 2018), we determined that the target sample size is 453.

We classified all participants according to their chronotype. To this end, we followed the guidelines put forward by Chelminski et al. (2000) and Griefahn et al. (2001). Accordingly, participants with values on the Morningness-Eveningness Questionnaire (D-MEQ, Griefahn et al., 2001) between 16 and 30 were categorized as definite evening types. Participants with values between 31 and 41 were categorized as moderate evening type. Participants with values between 42 and 58 were categorized as neutral type. Participants with values between 59 and 69 were categorized as moderate morning type, and participants with values between 70 and 86 were categorized as definite morning type. This categorization is the same as in the seminal research published by May and colleagues (see, e.g., Intons-Peterson et al., 1998; May, 1999; May et al., 1993).

In total, 689 young participants were tested. Participants were not paid. A complete description of the exclusion criteria used in the present study is presented in Table 2. Applying these criteria resulted in a sample consisting of 446 participants, which is close to our target sample size. Because previous research repeatedly reported a synchrony effect for participants categorized as moderate and definite

chronotypes (see, e.g., Intons-Peterson et al., 1998; May, 1999; May et al., 1993), we first report the analyses on the subsample including participants with moderate and definite chronotypes. The analyses on the full sample ($N = 446$) are presented as part of the multiverse-analysis approach.

The subsample consisted of 191 young adults. One hundred and thirty-one participants were morning types, and 60 were evening types. Demographic characteristics for this subsample are summarized in Table 3. The study was approved by the ethics committee of the university (approval number: 2020-06-00001), and all participants gave written informed consent.

Material

All tasks and questionnaires were programmed using lab.js (Henninger et al., 2019) on a computer using a 31 x 17.4 cm screen size and a 1920 x 1080 px screen resolution. A set of four tasks was used to measure short-term memory. These tasks were simple-span tasks with either digits, letters, matrices, or arrows as materials. They were programmed following Kane et al. (2004). A set of four tasks was used to measure working memory. These tasks were complex-span tasks and updating tasks with numerical or spatial materials. They were programmed following Rey-Mermet et al. (2019). Each task had two versions (i.e., Version 1 and Version 2). The two versions only differed in the presentation order of the stimulus exemplars. In each version, the same pseudorandom presentation order was administered for all participants. For all tasks, each stimulus exemplar was presented approximately equally often. Unless specified otherwise, each event (e.g., stimulus or prompt) was presented centrally in black color and in 36-point sans-serif font. In all tasks, the feedback consisted of a smiling face after a correct response and a frowning face after an error. Both feedback stimuli comprised a width and a height of 2.29° visual angle at a viewing distance of 60 cm. In the next paragraphs, we present the short-term memory and working-memory tasks as well as the questionnaires separately.

Short-term memory and working-memory tasks

In all short-term memory tasks, a trial consisted of an encoding phase followed by a recall phase. In the working-memory tasks, there was in addition either a distractor task for the complex spans or updating steps for the updating tasks. In all spans, set size refers to the number of memoranda to be remembered during each trial. Although the number of memoranda was limited in most tasks (e.g., the digits one until nine for the digit simple span), memoranda did not repeat within a trial. Next, we describe each task in detail.

In the digit simple span, the encoding phase consisted of memorizing sequences of digits. The digits were presented in set sizes ranging from two to nine digits. In the recall phase, digits had to be recalled in correct serial order. Thus, a text field was presented in the center of the screen. In addition, a counter (e.g., "digit 1" for the first digit to be re-

Table 2. Exclusion Criteria.

Reasons	Number of exclusions
Participants were not aged between 18 and 28.	9
Participants did not report Swiss German or German as native language.	5
Participants reported colorblindness or no normal vision.	0
Participants reported neurological or psychiatric disorders.	5
Participants did not complete the whole experiment.	46
The session order or the task order across the sessions was incorrect.	6
Participants took a rest longer than 30 minutes.	8
The data timestamp was not in chronological order.	2
Participants did not perform the morning session between 07:30 and 10:00 and the evening session between 16:30 and 19:00.	65
The morning or evening session lasted more than two hours.	2
A task was missing. ^a	69
Participants were multivariate outliers. ^b	26

^a A task was missing or discarded if the computer malfunctioned or if the participants did not respond within three minutes when performing the task. This was implemented to ensure laboratory-like setting.

^b We checked for multivariate normality across all measures using the Mardia's (1970) kurtosis index. Participants were considered as multivariate outliers when Mahalanobis's d^2 values were significant.

Table 3. Sample Characteristics and Background Measures.

Measure	Sample
Sample size	191
Age, years	23.8 (3)
Age range	18-28
Gender (female / male / other)	135 / 54 / 2
Education, years	13.5 (3.8)
Education level ^a	4.9 (1.5)
BDI-II score ^b	6.9 (6.3)
PSQ score ^c	30 (15.9)
D-MEQ chronoscore ^d (morning / evening)	64.5 (4) / 36.5 (4.1)
D-MEQ chronoscore range (morning / evening)	59-78 / 25-41

Note. Standard deviations are given in parentheses. BDI-II = Beck Depression Inventory II (Hautzinger et al., 2006); PSQ = Perceived Stress Questionnaire (Fliege et al., 2005); D-MEQ = German version of the Morningness-Eveningness-Questionnaire (Griefahn et al., 2001).

^a Education level ranged from 1 (no or less than 9 school years) to 8 (Ph.D.). ^b Depression score ranged from 0 (minimal depression) to 63 (severe depression). ^c PSQ score ranged from 0 to 100. ^d Chronoscore ranged from 16 (definite evening type) to 86 (definite morning type). Neutral types are indicated by a chronoscore ranging from 42 to 58.

called) was presented on the upper part of the screen (2.86° visual angle) to keep track of the serial order.

The letter simple span was similar to the digit simple span, except for the following modifications. First, the memoranda were the uppercase letters B, F, H, J, L, M, Q, R, and X. Second, set sizes ranged from two to eight letters. Third, at recall, letters had to be recalled in correct serial order by clicking sequentially on the corresponding letters in the matrix. To this end, all nine letters were presented in a 3 x 3 matrix (with a length of 4.68° visual angle). A counter (e.g., "letter 1" for the first letter to be recalled) was also presented in 32-point font on the upper part of the screen (3.82° visual angle).

The matrix simple span was similar to the letter simple span, except for the following modifications. First, the memoranda were the positions of squares in a 4 x 4 matrix. The matrix had a length of 4.96° visual angle and each

square of the matrix had a length of 1.24° visual angle. Second, set sizes ranged from two to seven squares. Third, at recall, an empty 4 x 4 matrix was displayed. The counter included the word "position" (e.g., "position 1" for the first position to be recalled).

The arrow simple span was similar to the letter and matrix simple spans, except for the following modifications. First, the memoranda were short and long arrows radiating out from the center of the screen. Short arrows had a length of 0.86° visual angle, whereas long arrows had a length of 1.72° visual angle. Each arrow pointed at either 0°, 45°, 90°, 135°, 180°, 225°, 270°, or 315°. Second, set size ranged from two to six arrows. Third, at recall, all short arrows were presented in a 3 x 3 matrix on the left side of the screen, and all long arrows were also presented in a 3 x 3 matrix but on the right side of the screen. Both matrices had a length of 6.30°, and they were separated by a length of 2.10° visual

angle. The counter included the word “arrow” (e.g., “arrow 1” for the first arrow to be recalled).

In the numerical complex span, the encoding phase consisted of memorizing sequences of three to five two-digit numbers. Between the presentation of these memoranda, the distractor task was presented. Thus, one equation that was either valid (e.g., “6-6=0”) or invalid (e.g., “5+11=18”) was displayed. The validity of the equation was judged by pressing the left- and right-pointing arrow key, respectively. This response mapping was presented in 24-point monospace font with a width of 12.35° visual angle on the lower part of the screen (4.30° visual angle). At recall, the numbers had to be recalled in correct serial order. Thus, a text field was presented in the center of the screen. In addition, a counter (e.g., “number 1” for the first number to be recalled) was presented on the upper part of the screen (2.86° visual angle) to keep track of the serial order.

The spatial complex span task was similar to the numerical complex span, except for the following modifications. First, four to six red squares were presented sequentially in a 5 x 5 matrix. The matrix had a length of 6.2° visual angle and each square of the matrix had a length of 1.24° visual angle. Second, the distractor task consisted of judging whether the pattern emerging from four squares presented concurrently and arranged in an L-shape was vertical or horizontal. Third, at recall, positions had to be recalled in correct serial order by clicking sequentially on the corresponding positions in the matrix. Thus, a 5 x 5 matrix was presented. In addition, a counter (e.g., “position 1” for the first position to be recalled) was displayed above the matrix (i.e., 4.30° visual angle).

In the numerical updating task, the encoding phase consisted of memorizing four digits (ranging from one to nine) presented in four different colors (i.e., red, blue, green, and orange). The digits were displayed centrally, each separated by a width of 1.43° visual angle. In the following updating steps, the digit to be updated was presented centrally in one of the four colors. In the recall phase, the most recent digit of each color had to be recalled. Thus, the word “digit” was displayed in the color corresponding to the digit to be recalled.

In the spatial updating task, the encoding phase consisted of memorizing the positions of three to five dots presented in a 4 x 4 matrix. The matrix had a length of 6.2° visual angle and each dot had a length of 1.24° visual angle. In each updating step, a new position of the to-be-updated dot was indicated by an arrow pointing in the direction of the required mental shift of that dot. Thus, one of the colored dots was presented centrally below a black arrow pointing either left, right, up, or down (with a width of 1.24° visual angle and a height of 0.48° visual angle). After each updating step, the most recent position of the dot to be updated had to be recalled. Thus, an empty 4 x 4 matrix is presented.

Questionnaires

In the present study, five questionnaires were used. The first questionnaire was the D-MEQ (Griefahn et al., 2001) to assess the chronotype of each participant. The second

questionnaire was implemented to have some descriptive information about our sample. Thus, this questionnaire assessed socio-demographic variables, such as age, gender, handedness, color blindness, nationality, native language, foreign language(s), number of education years, socio-economic status, synesthetic experience, and leisure activities (e.g., music, sport, video game) as well as information about the general health status and the preferences for dorsal/ventral processing types. The third questionnaire was used to have some information about the current health status of our sample. The questions were about medication use and sleep in the last 24 hours as well as nicotine, alcohol and drug consumption in the last 2 hours. Because of the known effects of depression and stress on cognition (e.g., McDermott & Ebmeier, 2009; Rock et al., 2013; Saenger et al., 2014; Starcke et al., 2016), the last two questionnaires were the German Version of the Beck Depression Inventory II (BDI-II, Hautzinger et al., 2006) and the German version of the Perceived Stress Questionnaire (PSQ, Fliege et al., 2005).

Procedure

Participants were tested remotely by means of a browser-based online experiment during three sessions. Participants were alone for all three sessions, but they could phone the student who recruited them in case of problems or questions. At the beginning of all three sessions, participants were required to confirm that they performed the experiment in a laboratory-like setting (e.g. they were alone in a quiet room, without distraction, and they closed all computer programs, but the browser window with the tasks). At the beginning of the second and third sessions, participants were asked to perform a scaling task. This task was implemented to ensure that all stimuli were presented with the same size because all tasks were run on participants' computers with different screen sizes and resolutions. In this scaling task, participants had to adapt the size of a rectangle to the size of a credit card. Then, based on this result, the size of the stimuli was computed so that it was the same across the different screen sizes and resolutions. In the middle of each session of the three sessions, participants could take a break of about 10 minutes. At the end of each session, they were asked about their current health status.

The three sessions were so organized that the first session was a screening session lasting for approximately 30 minutes. At the beginning of this session, after being informed, participants confirmed their consent to participate. Then, they were asked to perform the D-MEQ (Griefahn et al., 2001) in addition to a general socio-demographic questionnaire. The following two sessions lasted approximately for 1.5 hour each. These two sessions were separated by at least one night and maximally one week. The duration of one week was extended if the participant could not perform the session as planned (e.g., due to illness). The two sessions were performed either at 08:00 or at 17:00. We selected these testing times based on previous research (see Table 4). Moreover, because these two sessions lasted about 1.5 hour each, we considered testing times ranging from

07:30 to 10:00 and from 16:30 to 19:00, respectively, as still acceptable. These ranges are in line with the testing times reported in Table 4. For each chronotype (i.e., definite morning, moderate morning, neutral, moderate evening, and definite evening), half of the participants were assigned to be tested on peak in the second session and off peak in the third session, whereas the other half was assigned to be tested in the reversed order (i.e., off peak in the second session and on peak in the third session). During the on- and off-peak sessions, participants performed all short-term memory and working-memory tasks. At the end of the third session, participants were asked to complete the BDI-II (Hautzinger et al., 2006) and PSQ (Fliege et al., 2005).

Across the two on- and off- peaks sessions, the same order of short-term memory and working-memory tasks was used. For half of the participants, the following task order was used: verbal simple span, spatial complex span, numerical simple span, spatial updating, numerical complex span, arrow simple span, numerical updating, and spatial simple span. This task order was reversed for the other half of participants to control for practice effects. Moreover, in each session, one version of the tasks – that is, either Version 1 or Version 2 – was used. These versions and their order were counterbalanced across the two sessions so that half of the participants started with Version 1 and the other half with Version 2. Both counterbalancing conditions – that is, the counterbalancing of task order and of version order – were performed within each chronotype (i.e., within definite morning type, moderate morning type, neutral type, moderate evening type, and definite chronotype).

The task structure was similar across the different short-term memory and working-memory tasks. That is, each task started with the presentation of instructions explaining how participants had to carry out the task. These instructions were followed by a practice block, which could be repeated in case the participants required it. Following the practice block, participants performed one experimental block for the short-term memory tasks and two experimental blocks for the working-memory tasks. For the short-term memory tasks, the practice block included three trials with a set size of two, and the experimental block included three trials of each set size. In this experimental block, the set size ranged from three to nine for the digit simple span, three to eight for the letter simple span, two to seven for the matrix simple span, and two to six for the arrow simple span. For both numerical and spatial complex spans as well as for the spatial updating task, the practice block included two trials, and the two experimental blocks included a total of 12 trials. For the numerical updating task, the practice block included three trials (with four, six, and seven updating steps, respectively). The two experimental blocks included a total of 25 trials (each trial including seven updating steps). In addition, in this task, recall was probed in five out of the 25 trials immediately after the initial encoding. This was implemented to ensure that the initial set of memoranda was encoded. In all tasks, participants could take brief rests after each block.

The trial sequence was similar across all short-term memory tasks. That is, each trial started with the prompt

“Ready?” until the participant pressed the space key. Then, the memorandum was presented for 1000 ms in the digit, letter, and arrow simple spans, and for 650 ms in the matrix simple span. The presentation of each memorandum was followed by a blank screen for 500 ms. This sequence was repeated, depending on the set size. After all memoranda were presented, participants were required to recall the sequence of memoranda in correct serial order. In the digit simple span, participants were asked to enter their response with the keyboard and then to press “Next” on the screen using the mouse. In the letter, matrix, and arrow simple spans, they were asked to select their responses by clicking on the screen. During the practice block only, the feedback about the accuracy of the given response was presented for 500 ms. At the end of each trial, a blank screen was displayed for 500 ms.

Except for the prompt “Ready?” and the feedback during the practice block, the trial sequence was more diverse for the working-memory tasks. In the numerical and spatial complex spans, each memorandum was presented for 1000 ms, and each distractor task was presented for 3000 ms maximally. During the distractor task, the stimulus-response mapping was additionally presented but in the practice block only. This encoding sequence was repeated, depending on the set size. After all memoranda were presented, participants were asked to recall memoranda in correct serial order. In the numerical complex span, participants were asked to enter their response with the keyboard and then to press “Next” on the screen using the mouse. In the spatial complex span, they were asked to select their responses by clicking on the screen. In both complex spans, at the end of the trial, a blank screen was displayed for 500 ms. In the numerical updating task, the encoding phase lasted for 5000 ms, followed by a blank screen for 250 ms. Each updating step was then displayed for 1250 ms, followed by a blank screen for 250 ms. The recall lasted until the participant responded by entering the digit with the keyboard. In the spatial updating task, memoranda were simultaneously presented for 500 ms per colored dot (e.g., four memoranda were presented for 2000 ms). Each updating step lasted for 500 ms. After each updating step, participants were asked to recall the most recent position of the dot to be updated. To this end, the matrix was displayed again until the participant clicked on it to give a response.

Data preparation

For all short-term memory and working-memory tasks, the dependent measure consisted of the accuracy rates computed as the proportion of memoranda recalled at the correct position (partial-credit load score, see Conway et al., 2005). Mean accuracy rates were then computed for each participant, each task, and each session (on peak vs. off peak). For the numerical updating task, performance on the immediate probes were not included in the computation of the dependent measure. Standardized questionnaires were analyzed following their manual.

Table 4. Overview of the Testing Times Used in Previous Research.

Study	Morning testing time	Evening testing time
Bennett et al. (2008)	08:00-10:00	15:00-17:00
Bodenhausen (1990, Exp. 1)	09:00	20:00
Bodenhausen (1990, Exp. 2)	09:00	either 15:00 or 20:00
Ceglarek et al. (2021)	09:25-09:55; 11:00-11:30	18:30-19:00; 20:40-21:10
Fabbri et al. (2013, Exp. 1)	09:00-10:00	17:00-18:00
Fabbri et al. (2013, Exp. 2)	09:00-10:00	18:00-19:00
Goldstein et al. (2007)	8:00-10:00	13:00-15:00
Hahn et al. (2012)	8:00-10:00	13:00-15:00
Hasher et al. (2002)	8:00-9:15	16:30-17:15
Intons-Peterson et al. (1998)	8:00-10:30	15:30-18:00
Intons-Peterson et al. (1999, Exp. 1)	before 10:30	15:00 or later
Intons-Peterson et al. (1999, Exp. 3)	before 10:30	15:00 or later
Lara et al. (2014)	08:00	20:30
Lehmann et al. (2013)	9:00-11:00	15:00-17:00
Lewandowska et al. (2018)	8:00; 09:00	17:00; 18:00
Li et al. (1998, Exp. 1)	08:00	17:00
Li et al. (1998, Exp. 2)	08:00	17:00
Matchock and Mordkoff (2009)	08:00	16:00 and 20:00
May (1999)	08:00	17:00
May and Hasher (1998, Exp. 1)	08:00	16:00 or 17:00
May and Hasher (1998, Exp. 2)	08:00	17:00
May et al. (1993)	8:00 or 9:00	16:00 or 17:00
May et al. (2005, Exp. 1)	8:00-9:00	17:00-18:00
May et al. (2005, Exp. 2)	8:00-9:00	17:00-18:00
Petros et al. (1990)	09:00	20:00
Rothen and Meier (2016)	6:00-10:00	17:00-21:00
Rothen and Meier (2017)	8:00-12:00	16:00-20:00
Rowe et al. (2009)	8:00 or 9:00	16:00 or 17:00
Schmidt et al. (2012)	after one hour	after 10.5 hours
Schmidt et al. (2015)	after one hour	after 10.5 hours
Van Opstaal (2021)	08:00	20:30
West et al. (2002)	09:00	17:00
Yang et al. (2007, Exp. 1)	9:00-10:00	16:00-17:00
Yang et al. (2007, Exp. 2)	9:00-10:00	16:00-17:00
Yaremenko et al. (2021)	07:40-09:00	20:30-21:30
Yoon (1997)	8:00 or 9:00	16:00 or 17:00

Note. For Ceglarek et al. (2021) and Lewandowska et al. (2018), the testing times depended on whether the participants were categorized as morning or evening types. In each cell, the first testing time referred to the testing time used for the morning chronotype; the second testing time referred to the testing times used for the evening chronotype. For Schmidt et al. (2012, 2015), the testing times were individually selected so that it was either after one hour of wakefulness or after 10.5 hours.

Data analysis

We used an alpha level of .05 for all tests from the null hypothesis significance testing (NHST) framework. Effect sizes were calculated as Cohen's *d*. For the Bayesian hypothesis testing, default prior scales were used. Moreover, the Bayes Factors (BFs) were interpreted using Raftery's (1995) classification scheme. According to this classification, a BF between 1-3 is considered as weak evidence, a BF between 3-20 is considered as positive evidence, a BF

between 20-150 is considered as strong evidence, and a BF larger than 150 is considered as very strong evidence.

Model estimation

Model fit was evaluated via multiple fit indices (Hu & Bentler, 1998, 1999): the χ^2 goodness-of-fit statistic, the Bentler's comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root-mean-square residual (SRMR). For the χ^2 statistic, a small, non-significant value indicates good fit. For the CFI,

values larger than .95 indicate good fit, and values between .90 and .95 indicate acceptable fit. RMSEA values smaller than .06 and SRMR values smaller than .08 indicate good fit. It is to note that RMSEA values are less preferable when the sample size includes less than 250 participants (Hu & Bentler, 1998). In such cases, these values are provided for the sake of completeness, but they are not taken into account for evaluating the model fit.

In addition, the following criteria had to be met for a model to be considered a “good” fitting model: (1) the Kaiser-Meyer-Olkin (KMO) index – a measure of whether the correlation matrix is factorable – should be larger than .60 (Tabachnick & Fidell, 2019); (2) most of the error variances had to be lower than .90; (3) most of the factor loadings had to be significant and larger than .30; (4) no factor should be dominated by a large loading from one task; (5) The quality of how well the factor was represented by a set of measures – a measure sometimes called construct reliability or replicability – had to be good. This was assessed by the index H and it had to meet the standard criterion of .70 (Rodriguez et al., 2016).

Results

Results are reported in three steps. First, we investigated the reliability estimates and the correlational pattern for all measures. Second, we replicated previous research by examining the synchrony effect at the individual-task level (see, e.g., Ceglarek et al., 2021; Lewandowska et al., 2017; Rowe et al., 2009; Schmidt et al., 2015; West et al., 2002). Third, we used structural equation modeling (SEM) to investigate the synchrony effect at the latent-variable level for the constructs of short-term maintenance and attentional control. In this step, we aimed to model these constructs for on and off peaks separately and to estimate a latent-change model between both peaks. This assessment is equivalent to a paired t -test which is applied to latent constructs (see Kievit et al., 2018).

Reliability and correlations

As shown in Table 5, all measures had acceptable skew and kurtosis (i.e., between -1.01 and 0.50). The reliability estimates for all measures in both sessions were good, ranging from .77 to .95.

Pearson correlation coefficients as well as their upper and lower confidence intervals are shown in Table 6. Bayes factors (BFs) for the correlations are presented in Table 7. These assessed the weight of evidence in favor of the alternative hypothesis (BF_{10} , i.e., in favor of a correlation) and in favor of the null hypothesis (BF_{01} , i.e., in favor of the absence of the correlation). The correlations were moderate to strong, ranging from .24 to .73. All correlations were significant ($ps < .001$), and all Bayes Factors (BFs) suggested strong to very strong evidence for the correlations (all $BF_{10} \geq 38.92$).

Results at the individual-task level: Synchrony effect in short-term memory and working-memory tasks

Consistent with previous research (e.g., Ceglarek et al., 2021; Lewandowska et al., 2017; Rowe et al., 2009; Schmidt et al., 2015; West et al., 2002), we then investigated the synchrony effect at the individual-task level. This is displayed in Figure 1. Statistically, we compared short-term memory and working-memory performance between on- and off-peak sessions by computing a paired two-tailed t -test for each short-term memory and working-memory task separately. In addition, we computed Bayesian t -tests. This allowed us to assess not only the strength of evidence for the alternative hypothesis (i.e., the presence of the synchrony effect) but also the strength of evidence for the null hypothesis (i.e., the absence of a synchrony effect). These results are presented in Table 8. As shown in this table, no synchrony effect was observed for most tasks. Only in the numerical updating task and in the arrow simple span, a significant synchrony effect was found. However, the effect sizes were small for these tasks, and Bayesian evidence only suggests weak to positive evidence in favor of the effect (see Table 8).

Results at the latent-variable level: Synchrony effect in short-term maintenance and attentional control

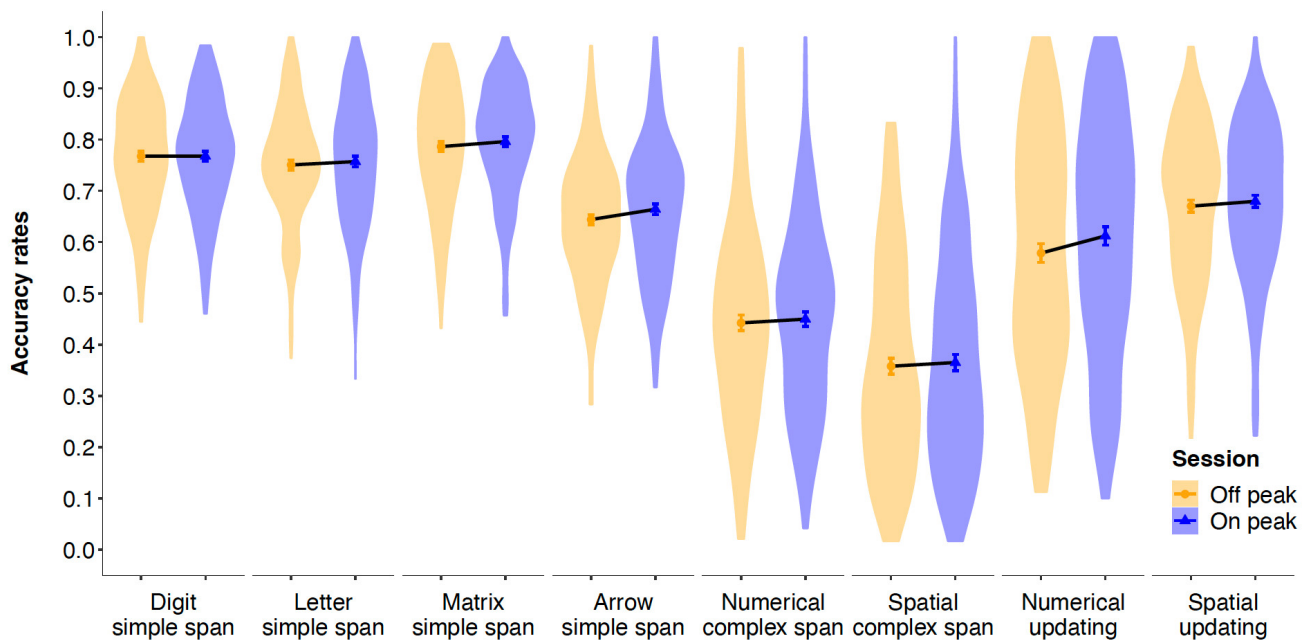
The synchrony effect in short-term maintenance and attentional control was assessed by modeling a latent change between both on and off peaks for these constructs. A bi-factor model was used to model short-term maintenance and attentional control. Overall, four latent-change models were thus modeled (see Figures 2 and 3). In each model, we took into account measure-specific variance across both sessions by allowing error variances from the measures of the same task to correlate. Factor loadings were also constrained to be positive. Moreover, the measurement invariance across time was modeled by applying equality constraints over off and on peaks for the factor loadings, the error variances, and the intercepts. Furthermore, we assessed the synchrony effect for each latent construct by modeling a latent change between the on-peak factor and the off-peak factor of the construct. That is, the on-peak factor was regressed on the off-peak factor by fixing the unstandardized regression weight to 1. The latent-change factor Δ was then measured by fixing the unstandardized factor loading to the on-peak factor to 1. A covariance was also added between the off-peak factor and the latent-change factor Δ to capture whether the latent change was dependent or proportional to the scores at off peak.

Following this strategy, we estimated the first model (Model 1) by modeling short-term maintenance as the common variance between short-term memory and working-memory measures, and attentional control as the working-memory variance remaining after controlling for short-term memory (see Engle et al., 1999). More precisely, all short-term memory and working-memory measures at off and on peaks were forced to load on an off-peak main-

Table 5. Synchrony Effect: Descriptive Statistics.

Session	Task	Mean	SD	Min.	Max.	Skew	Kurtosis	Reliability
Off peak	Digit simple span	.77	.11	.44	1	-0.36	-0.23	.89 [.87, .91]
	Letter simple span	.75	.12	.37	1	-0.46	0.37	.87 [.84, .89]
	Matrix simple span	.79	.12	.43	.99	-0.49	-0.32	.87 [.84, .89]
	Arrow simple span	.64	.12	.28	.98	-0.13	0.19	.77 [.72, .81]
	Numerical complex span	.44	.19	.02	.98	0.28	-0.01	.88 [.85, .90]
	Spatial complex span	.36	.20	.02	.83	0.48	-0.70	.92 [.91, .94]
	Numerical updating	.58	.23	.11	1	-0.05	-1.01	.95 [.94, .96]
	Spatial updating	.67	.16	.11	.98	-0.64	0.30	.93 [.91, .94]
On peak	Digit simple span	.77	.11	.46	.98	-0.32	-0.32	.89 [.87, .91]
	Letter simple span	.76	.12	.33	1	-0.40	-0.03	.88 [.86, .91]
	Matrix simple span	.80	.11	.46	1	-0.75	0.33	.84 [.81, .88]
	Arrow simple span	.66	.13	.32	1	-0.07	-0.20	.80 [.76, .84]
	Numerical complex span	.45	.18	.04	1	0.21	-0.20	.86 [.84, .89]
	Spatial complex span	.37	.20	.02	1	0.50	-0.39	.92 [.91, .94]
	Numerical updating	.61	.23	.10	1	-0.24	-0.89	.95 [.94, .96]
	Spatial updating	.68	.15	.22	1	-0.65	0.30	.92 [.90, .93]

Note. Short-term memory and working memory were measured using accuracy rates. Permutation-based split-half reliability estimates were computed (see Parsons et al., 2019). The split-half correlations were adjusted with the Spearman–Brown prophecy formula, and the results of 5000 random splits were averaged. The 95% confidence intervals are presented in brackets. SD = Standard Deviation; Min. = minimum; Max. = maximum.

**Figure 1. Synchrony Effect (i.e., Mean Accuracy Difference between On- and Off-Peak Sessions) for Each Short-Term and Working-Memory Task**

Note. Error bars represent within-subject confidence intervals (Cousineau, 2005; Morey, 2008). Shaded areas display data density plots.

tenance factor and an on-peak maintenance factor, respectively. In addition, all working-memory tasks at off and on peaks were forced to load on an off-peak attentional-control factor and an on-peak attentional-control factor, respectively. We assessed the synchrony effect for maintenance by modeling a latent change between the on-peak maintenance factor and the off-peak maintenance factor.

Similarly, the synchrony effect for attentional control was assessed by modeling a latent change between the on-peak attentional-control factor and the off-peak attentional-control factor. This model is depicted in Figure 2a. However, it provided a bad fit to the data, $KMO = .86$, $\chi^2(97, N = 191) = 266.44$, $p < .001$, $CFI = .91$, $RMSEA [90\% CI] = .10 [.08, .11]$, $SRMR = .09$.

Table 6. Synchrony Effect: Pearson Correlation Coefficients.

Session	Task	Off peak										On peak				
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Off peak	1. Digit s.s.	-														
	2. Letter s.s.	.68* [.60, .75]	-													
	3. Matrix s.s.	.26* [.12, .39]	.28* [.14, .40]	-												
	4. Arrow s.s.	.37* [.24, .48]	.33* [.19, .45]	.59* [.49, .68]	-											
	5. Numerical c.s.	.54* [.43, .64]	.54* [.43, .63]	.34* [.20, .46]	.40* [.27, .51]	-										
	6. Spatial c.s.	.32* [.18, .44]	.35* [.22, .47]	.55* [.44, .64]	.40* [.27, .51]	.45* [.33, .56]	-									
	7. Numerical upd.	.34* [.21, .46]	.36* [.23, .48]	.39* [.27, .51]	.47* [.35, .57]	.40* [.28, .52]	.43* [.31, .54]	-								
	8. Spatial upd.	.30* [.17, .43]	.29* [.15, .41]	.56* [.45, .65]	.53* [.41, .62]	.34* [.21, .46]	.50* [.38, .60]	.50* [.39, .60]	-							
On peak	9. Digit s.s.	.67* [.59, .75]	.63* [.53, .71]	.32* [.19, .44]	.38* [.25, .49]	.44* [.32, .55]	.36* [.23, .48]	.28* [.15, .41]	.25* [.12, .38]	-						
	10. Letter s.s.	.60* [.50, .68]	.68* [.60, .75]	.27* [.13, .39]	.32* [.18, .44]	.49* [.37, .59]	.29* [.15, .41]	.25* [.11, .38]	.27* [.13, .40]	.66* [.57, .73]	-					
	11. Matrix s.s.	.28* [.14, .41]	.34* [.21, .46]	.70* [.62, .77]	.52* [.41, .62]	.25* [.11, .38]	.42* [.30, .53]	.25* [.12, .38]	.45* [.33, .56]	.38* [.26, .50]	.31* [.18, .44]	-				
	12. Arrow s.s.	.29* [.15, .41]	.26* [.12, .38]	.58* [.47, .66]	.70* [.61, .76]	.31* [.18, .43]	.38* [.25, .50]	.39* [.26, .50]	.49* [.37, .59]	.41* [.28, .52]	.32* [.18, .44]	.55* [.44, .64]	-			
	13. Numerical c.s.	.44* [.32, .55]	.53* [.42, .63]	.31* [.17, .43]	.34* [.21, .46]	.67* [.58, .74]	.35* [.22, .47]	.37* [.24, .49]	.24* [.10, .37]	.52* [.41, .62]	.52* [.41, .62]	.34* [.21, .46]	.42* [.29, .53]	-		
	14. Spatial c.s.	.26* [.13, .39]	.31* [.17, .43]	.51* [.40, .61]	.32* [.19, .44]	.32* [.18, .44]	.72* [.64, .78]	.29* [.16, .42]	.45* [.33, .56]	.42* [.30, .53]	.39* [.26, .50]	.46* [.34, .56]	.39* [.26, .50]	.43* [.31, .54]	-	
	15. Numerical upd.	.28* [.14, .40]	.35* [.22, .47]	.33* [.20, .45]	.34* [.21, .46]	.31* [.18, .44]	.40* [.27, .51]	.64* [.49, .71]	.41* [.29, .53]	.37* [.24, .48]	.33* [.20, .45]	.36* [.23, .48]	.42* [.29, .53]	.50* [.39, .60]	.44* [.32, .55]	-
	16. Spatial upd.	.27* [.14, .40]	.28* [.15, .41]	.48* [.36, .58]	.44* [.32, .55]	.28* [.14, .40]	.41* [.28, .52]	.33* [.20, .46]	.73* [.65, .79]	.30* [.17, .42]	.35* [.22, .47]	.51* [.39, .61]	.46* [.34, .56]	.29* [.15, .41]	.46* [.34, .56]	.42* [.30, .53]

Note. Ninety-five percent confidence intervals are presented in brackets. Correlations for which the Bayes factor suggested positive to strong evidence for the alternative hypothesis (BF_{10}) are presented in bold; correlations for which the Bayes factor suggested positive to strong evidence for the null hypothesis (BF_{01}) are presented in italics. S.s. = simple span; c.s. = complex span; upd. = updating.

* $p < .05$.

Table 7. Synchrony Effect: Bayes Factors in Favor of the Alternative Hypothesis (BF₁₀) and in Favor of the Null Hypothesis (BF₀₁) for the Pearson Correlation Coefficients.

Session	Task	BF	Off peak										On peak				
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Off peak	1. Digit s.s.	BF ₁₀	-														
		BF ₀₁															
	2. Letter s.s.	BF ₁₀	1.60*10 ²⁴	-													
		BF ₀₁	6.24*10 ⁻²⁵														
	3. Matrix s.s.	BF ₁₀	98.48	278.69	-												
		BF ₀₁	0.01	3.59*10 ⁻³													
	4. Arrow s.s.	BF ₁₀	8.46*10 ⁴	5.32*10 ³	2.27*10 ¹⁶	-											
		BF ₀₁	1.18*10 ⁻⁵	1.88*10 ⁻⁴	4.40*10 ⁻¹⁷												
	5. Numerical c.s.	BF ₁₀	1.34*10 ¹³	6.32*10 ¹²	9.83*10 ³	1.19*10 ⁶	-										
		BF ₀₁	7.48*10 ⁻¹⁴	1.58*10 ⁻¹³	1.02*10 ⁻⁴	8.37*10 ⁻⁷											
	6. Spatial c.s.	BF ₁₀	2.74*10 ³	2.88*10 ⁴	2.79*10 ¹³	1.40*10 ⁶	2.10*10 ⁸	-									
		BF ₀₁	3.64*10 ⁻⁴	3.47*10 ⁻⁵	3.58*10 ⁻¹⁴	7.16*10 ⁻⁷	4.76*10 ⁻⁹										
	7. Numerical upd.	BF ₁₀	1.65*10 ⁴	5.62*10 ⁴	8.10*10 ⁵	1.21*10 ⁹	2.06*10 ⁶	2.60*10 ⁷	-								
		BF ₀₁	6.05*10 ⁻⁵	1.78*10 ⁻⁵	1.23*10 ⁻⁶	8.29*10 ⁻¹⁰	4.85*10 ⁻⁷	3.84*10 ⁻⁸									
	8. Spatial upd.	BF ₁₀	1.08*10 ³	523.30	7.90*10 ¹³	1.28*10 ¹²	1.24*10 ⁴	3.99*10 ¹⁰	4.95*10 ¹⁰	-							
		BF ₀₁	9.25*10 ⁻⁴	1.91*10 ⁻³	1.27*10 ⁻¹⁴	7.83*10 ⁻¹³	8.07*10 ⁻⁵	2.51*10 ⁻¹¹	2.02*10 ⁻¹¹								
On peak	9. Digit s.s.	BF ₁₀	2.44*10 ²³	1.45*10 ¹⁹	3.27*10 ³	1.89*10 ⁵	1.01*10 ⁸	5.62*10 ⁴	360.76	76.89	-						
		BF ₀₁	4.11*10 ⁻²⁴	6.91*10 ⁻²⁰	3.06*10 ⁻⁴	5.28*10 ⁻⁶	9.92*10 ⁻⁹	1.78*10 ⁻⁵	2.77*10 ⁻³	0.01							
	10. Letter s.s.	BF ₁₀	8.73*10 ¹⁶	1.87*10 ²⁴	135.18	2.56*10 ³	1.09*10 ¹⁰	510.81	68.38	188.91	6.20*10 ²¹	-					
		BF ₀₁	1.15*10 ⁻¹⁷	5.35*10 ⁻²⁵	7.40*10 ⁻³	3.90*10 ⁻⁴	9.15*10 ⁻¹¹	1.96*10 ⁻³	0.01	5.29*10 ⁻³	1.61*10 ⁻²²						
	11. Matrix s.s.	BF ₁₀	300.22	1.31*10 ⁴	2.42*10 ²⁶	1.02*10 ¹²	62.66	1.09*10 ⁷	78.89	1.45*10 ⁸	3.68*10 ⁵	2.28*10 ³	-				
		BF ₀₁	3.33*10 ⁻³	7.61*10 ⁻⁵	4.14*10 ⁻²⁷	9.81*10 ⁻¹³	0.02	9.21*10 ⁻⁸	0.01	6.88*10 ⁻⁹	2.72*10 ⁻⁶	4.38*10 ⁻⁴					
	12. Arrow s.s.	BF ₁₀	456.42	81.46	2.21*10 ¹⁵	3.02*10 ²⁵	1.92*10 ³	3.11*10 ⁵	4.15*10 ⁵	1.08*10 ¹⁰	3.05*10 ⁶	2.45*10 ³	4.09*10 ¹³	-			
		BF ₀₁	2.19*10 ⁻³	0.01	4.53*10 ⁻¹⁶	3.31*10 ⁻²⁶	5.22*10 ⁻⁴	3.22*10 ⁻⁶	2.41*10 ⁻⁶	9.29*10 ⁻¹¹	3.28*10 ⁻⁷	4.08*10 ⁻⁴	2.44*10 ⁻¹⁴				
	13. Numerical c.s.	BF ₁₀	5.19*10 ⁷	3.20*10 ¹²	1.66*10 ³	1.41*10 ⁴	3.47*10 ²²	2.19*10 ⁴	1.10*10 ⁵	38.92	9.85*10 ¹¹	7.80*10 ¹¹	1.67*10 ⁴	5.92*10 ⁶	-		
		BF ₀₁	1.93*10 ⁻⁸	3.13*10 ⁻¹³	6.04*10 ⁻⁴	7.08*10 ⁻⁵	2.88*10 ⁻²³	4.58*10 ⁻⁵	9.08*10 ⁻⁶	0.03	1.02*10 ⁻¹²	1.28*10 ⁻¹²	5.98*10 ⁻⁵	1.69*10 ⁻⁷			
	14. Spatial c.s.	BF ₁₀	128.81	1.43*10 ³	2.25*10 ¹¹	3.54*10 ³	2.83*10 ³	1.10*10 ²⁸	586.53	1.84*10 ⁸	1.26*10 ⁷	5.34*10 ⁵	4.78*10 ⁸	4.13*10 ⁵	2.97*10 ⁷	-	
		BF ₀₁	7.76*10 ⁻³	6.99*10 ⁻⁴	4.45*10 ⁻¹²	2.83*10 ⁻⁴	3.53*10 ⁻⁴	9.13*10 ⁻²⁹	1.70*10 ⁻³	5.43*10 ⁻⁹	7.93*10 ⁻⁸	1.87*10 ⁻⁶	2.09*10 ⁻⁹	2.42*10 ⁻⁶	3.36*10 ⁻⁸		
	15. Numerical upd.	BF ₁₀	274.86	2.55*10 ⁴	6.34*10 ³	1.20*10 ⁴	2.25*10 ³	1.30*10 ⁶	9.09*10 ¹⁹	5.23*10 ⁶	9.16*10 ⁴	5.93*10 ³	5.68*10 ⁴	6.39*10 ⁶	6.40*10 ¹⁰	6.35*10 ⁷	-
		BF ₀₁	3.64*10 ⁻³	3.92*10 ⁻⁵	1.58*10 ⁻⁴	8.36*10 ⁻⁵	4.45*10 ⁻⁴	7.71*10 ⁻⁷	1.10*10 ⁻²⁰	1.91*10 ⁻⁷	1.09*10 ⁻⁵	1.69*10 ⁻⁴	1.76*10 ⁻⁵	1.56*10 ⁻⁷	1.56*10 ⁻¹¹	1.57*10 ⁻⁸	
	16. Spatial upd.	BF ₁₀	225.33	335.53	2.66*10 ⁹	5.67*10 ⁷	267.41	3.02*10 ⁶	9.14*10 ³	1.85*10 ²⁹	959.14	3.78*10 ⁴	1.24*10 ¹¹	3.55*10 ⁸	467.77	4.35*10 ⁸	1.17*10 ⁷
		BF ₀₁	4.44*10 ⁻³	2.98*10 ⁻³	3.76*10 ⁻¹⁰	1.76*10 ⁻⁸	3.74*10 ⁻³	3.31*10 ⁻⁷	1.09*10 ⁻⁴	5.41*10 ⁻³⁰	1.04*10 ⁻³	2.65*10 ⁻⁵	8.04*10 ⁻¹²	2.81*10 ⁻⁹	2.14*10 ⁻³	2.30*10 ⁻⁹	8.57*10 ⁻⁸

Note. S.s. = simple span; c.s. = complex span; upd. = updating.

Table 8. Synchrony Effect: Inferential Statistical Values for the *t*-Tests Comparing On- and Off-Peak Sessions for Each Short-Term Memory and Working-Memory Task Separately, and Bayes Factors (BF) from the Bayesian *t*-Tests.

Task	<i>t</i> (190)	<i>p</i>	Cohen's <i>d</i>	BF ₁₀	BF ₀₁
<i>Digit simple span</i>	0.01	.988	0.00	0.08	12.37
<i>Letter simple span</i>	0.96	.338	0.07	0.13	7.86
<i>Matrix simple span</i>	1.53	.128	0.11	0.25	3.93
Arrow simple span	2.75	.007*	0.20	3.13	0.32
<i>Numerical complex span</i>	0.66	.509	0.05	0.10	9.98
<i>Spatial complex span</i>	0.65	.514	0.05	0.10	10.03
Numerical updating	2.38	.018*	0.17	1.27	0.79
<i>Spatial updating</i>	1.15	.252	0.08	0.15	6.47

Note. For the sake of clarity, results with a BF₁₀ larger than 3 are presented in bold, whereas results with a BF₀₁ larger than 3 are presented in italics.

* $p < .05$.

Short-term memory and working-memory tasks were reported to involve different maintenance processes for verbal-numerical and spatial materials, but common attentional-control processes (see, e.g., Kane et al., 2004). Accordingly, we fitted a second model (Model 2), which was similar to Model 1, except that the off- and on-peak maintenance factors were modeled separately for both material types. This model is depicted in Figure 2b. It provided a good fit to the data, $KMO = .86$, $\chi^2(87, N = 191) = 133.68$, $p = .001$, $CFI = .97$, $RMSEA [90\% CI] = .05 [.03, .07]$, $SRMR = .07$. However, the model parameters were not fully identified as the covariance matrix was not positive definite.

Applying a model-reduction strategy, we fitted a third model (Model 3) in which we assumed no synchrony effect for short-term maintenance. The synchrony effect was modeled only for attentional control. Thus, this model was similar to Model 2, except that there was one maintenance factor for verbal-numerical material and one maintenance factor for spatial material across both on- and off-peak sessions. This model is depicted in Figure 3a. It provided an acceptable fit to the data, $KMO = .86$, $\chi^2(101, N = 191) = 210.16$, $p < .001$, $CFI = .94$, $RMSEA [90\% CI] = .08 [.06, .09]$, $SRMR = .08$. Inspection of the key parameter for the latent change shows no difference between off and on peaks, thus indicating no synchrony effect for attentional control (the unstandardized intercept of the latent-change factor = .01, $se = .01$, $p = .493$, $95\% CI = [-.01, .03]$). In this model, however, the construct reliability for the attentional-control factors was low ($H = .42$ and $H = .39$ for off and on peak, respectively). This indicates that the attentional-control factors did not represent much common variance across the measures. Therefore, this model has low explanatory power.

In the last model (Model 4), we followed previous work in which attentional control has been modeled as a common factor across all short-term memory and working-memory measures (see Kane et al., 2004). Thus, short-term memory and working-memory tasks at off and on peaks were forced to load on an off-peak general factor and an on-peak general factor, respectively. In addition, short-term memory and working-memory tasks with verbal-numerical ma-

terial were forced to load on a verbal-numerical maintenance factor, whereas those tasks with spatial material were forced to load on a spatial maintenance factor. Thus, similar to Model 3, the synchrony effect was modeled only for attentional control. This model is depicted in Figure 3b. It provided a good fit to the data, $KMO = .86$, $\chi^2(94, N = 191) = 110.49$, $p = .118$, $CFI = .99$, $RMSEA [90\% CI] = .03 [.00, .05]$, $SRMR = .05$. In this case, the construct reliability for all factors was good (all $Hs \geq .73$). Inspection of the key parameter for the latent change shows no difference between off and on peaks, thus indicating no synchrony effect for attentional control (the unstandardized intercept of the latent-change factor = .001, $se = .01$, $p = .937$, $95\% CI = [-.01, .01]$).

Multiverse-analysis approach

In a final step, we tested for the robustness of the results in a multiverse-analysis approach. To this end, we re-ran the analyses by applying different data transformations, participants' selections, trimming procedures, and SEM approaches. The different procedures we used are described in Table 9. Please note that as preregistered, we modeled the data using a second-order latent growth curve modeling approach. However, these model assessments never resulted in acceptable or good fit statistics. Therefore, we opted for the latent-change models, which is an extension of the latent growth curve models focusing more directly on the difference between the two measurement time points (Ghisletta & McArdle, 2012).

An overview of the results from the multiverse-analysis approach is presented in Figure 4. As shown in the upper part of the figure, no robust synchrony effect was observed at the individual-task level. At the latent-variable level, the results showed the difficulty of measuring short-term maintenance and attentional control as good factors. When these were modeled satisfactorily (i.e., with acceptable to good fit statistics and good indices H), the synchrony effect was in most cases not significant. In the two cases showing a significant effect, the latent change was small (maximum value for the unstandardized intercept of the latent-change factor = 0.02).

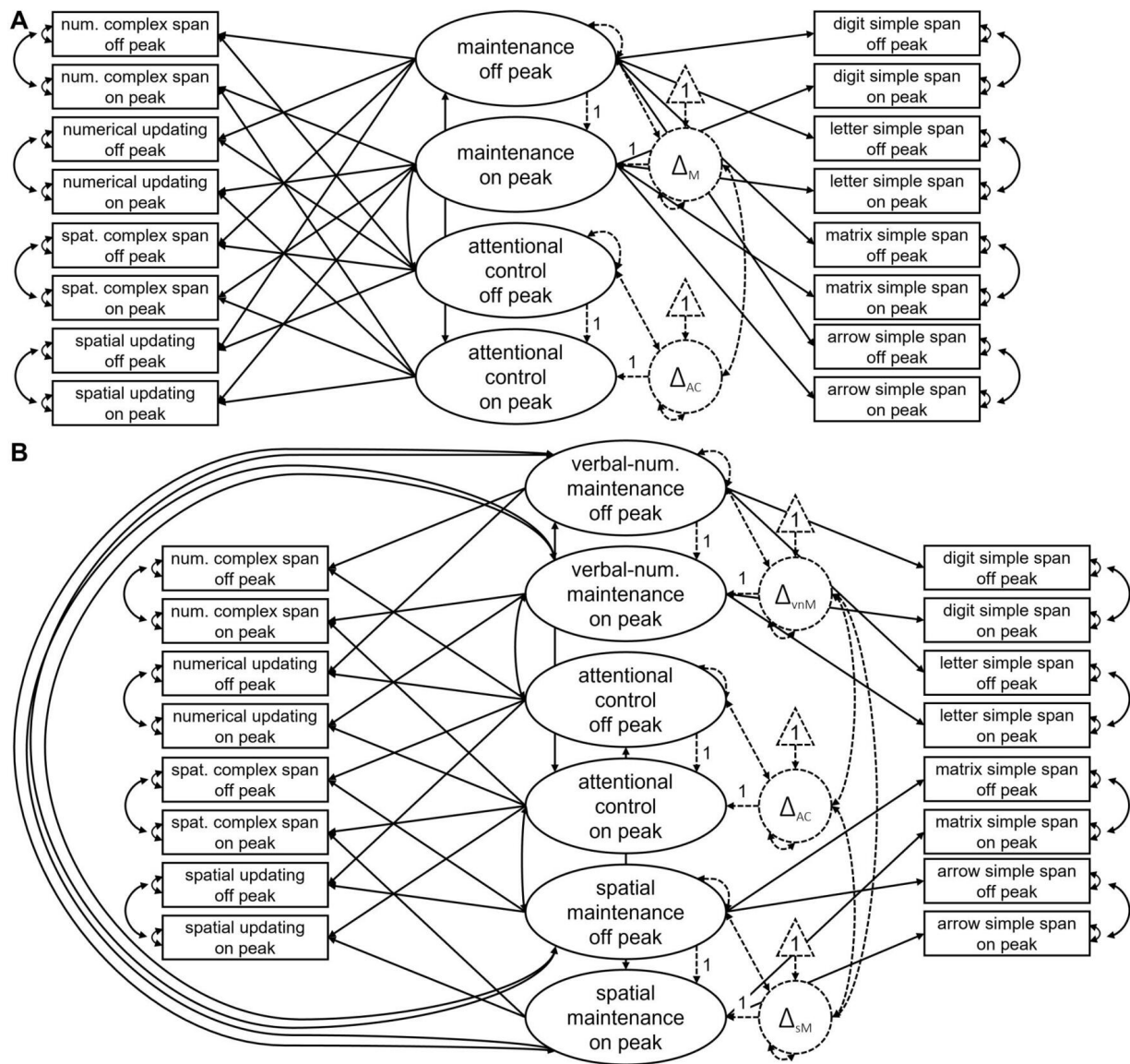


Figure 2. Latent-Change Model 1 and Latent-Change Model 2

Note. Panel A: Model 1. Panel B: Model 2. Please see the text for a detailed description of the model parameters. In both models, the synchrony effect was estimated as the latent change between off and on peak for each latent construct. Paths and variables specific to the estimation of the latent changes are displayed in dashed lines. Means are omitted for visual clarity. Num. = numerical; spat. = spatial; Δ_{AC} = latent-change factor for attentional control; Δ_M = latent-change factor for short-term maintenance; Δ_{vnM} = latent-change factor for verbal-numerical short-term maintenance; Δ_{sM} = latent-change factor for spatial short-term maintenance.

Finally, for the sake of completeness and following previous work (e.g., Allen et al., 2008; Bonnefond et al., 2003; Fabbri et al., 2008; Lewandowska et al., 2017; Matchock & Mordkoff, 2008), we also investigated the time-of-day effect, that is, the difference in performance between the morning and evening sessions (but independently of the subjective peak of circadian arousal). The results are summarized in the Appendix B. Consistent with the results about the synchrony effect, the results showed no robust time-of-day effect either at the individual-task level or at the latent-variable level.

Discussion

Most individuals who are classified as morning types according to their morningness-eveningness preferences have their peak of circadian arousal in the morning. It is assumed

that they show better cognitive performance in the morning than in the evening. Conversely, most individuals who are classified as evening types have their peak of circadian arousal in the evening. It is assumed that they show better cognitive performance in the evening than in the morning. This synchrony effect – that is, the observation of better cognitive performance at the peak of circadian arousal than at off peak – is well-established as common knowledge. However, empirical evidence is more equivocal. In the present study, we aimed to empirically clarify this effect. Specifically, we determined the scope and robustness of the synchrony effect by addressing the methodical challenges typically observed in previous research. Thus, we investigated the synchrony effect on short-term memory, working memory, and attentional control in a large sample of participants, who were tested at their on-peak time and their off-peak time. Following seminal research (see, e.g.,

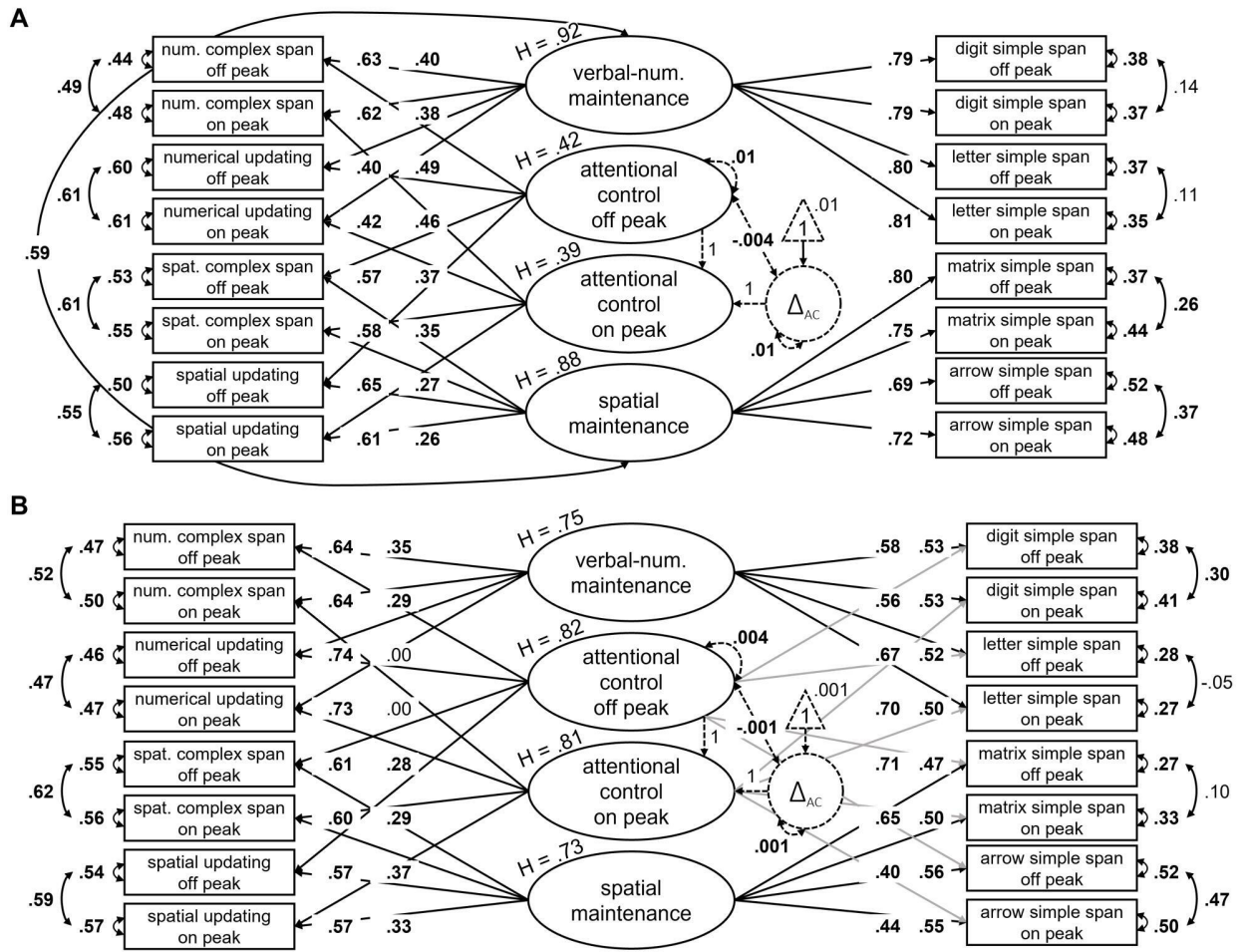


Figure 3. Latent-Change Model 3 and Latent-Change Model 4

Note. Panel A: Model 3. Panel B: Model 4. See Figure 2 note for details. The numbers next to the straight, single-headed arrows are the standardized factor loadings (interpretable as standardized regression coefficients). For the sake of clarity, the loadings are aligned to each measure (i.e., next to the measures for the general factor(s) and next to the factors for the specific factor(s)). The outward numbers adjacent to each measure are the error variances, attributable to idiosyncratic task requirements and measurement error. The numbers adjacent to the curved, double-headed arrows next to each measure are the correlations between the error variances. Parameters specific to the latent change are unstandardized. For all values, boldface type indicates $p < .05$. Num. = numerical; spat. = spatial; Δ_{AC} = latent-change factor for attentional control.

Intons-Peterson et al., 1998; May, 1999; May et al., 1993), on- and off-peak times were determined using a questionnaire (Griefahn et al., 2001; Horne & Östberg, 1976). All participants performed four short-term memory tasks and four working-memory tasks. Attentional control was assessed on the latent-variable level as the goal-directed nature of working-memory tasks without their maintenance aspects. Quite surprisingly, the results showed no evidence for a general and robust synchrony effect for any of the constructs we measured (i.e., short-term memory, working memory, and attentional control). Moreover, this pattern of results was confirmed when we applied different data transformations, participants' selection criteria, and trimming procedures.

On the individual-task level, the results showed no synchrony effect for most tasks. A synchrony effect was only detected in one out of four short-term memory tasks (i.e., arrow simple span) and one out of four working-memory tasks (i.e., numerical updating task). However, their effect sizes were at best small. From a positive perspective, our results can be interpreted that the synchrony effect does exist and that it would emerge in about 25% of cognitive tasks

(i.e., 2/8 of our tasks). From this perspective and based on the assumption that a large portion of null-findings are not published (e.g., Kühberger et al., 2014), our findings would be broadly consistent with previous research showing mixed evidence with regards to the synchrony effect on short-term memory and working memory (e.g., Rowe et al., 2009; Schmidt et al., 2015; West et al., 2002; but Ceglarek et al., 2021; Heimola et al., 2021; Lewandowska et al., 2017). This interpretation of the findings would be valid if the circumstances under which a synchrony effect occurs and under which it does not occur were clear. For example, the circumstances would have been clear if the synchrony effect was observed in one task type (e.g., working-memory tasks) or – although theoretically less plausible – only in tasks with the same stimulus materials (e.g., the digit simple span in short-term memory and the numerical complex span in working memory). However, we observed a synchrony effect in two tasks (arrow simple span and numerical updating) with different types (short-term memory and working memory, respectively) and different materials (spatial and numerical, respectively). Therefore, these results indicate no systematic pattern. Given that, our posi-

Table 9. Multiverse-analysis approach: Description of the Different Procedures Used in the Present Study.

Procedure	Description
Data transformation	1. raw accuracy rates 2. arcus-sinus transformed accuracy rates
Participants' selection	1. all chronotypes ^a 2. moderate and definite morning and evening types 3. definite morning and evening types
Trimming	1. with missing data (but no more than 2 short-term or working-memory tasks were missing in each session) 2. after removing missing data 3. after removing outliers in time-out responses (i.e., number of time-out responses smaller or larger than 3 SDs) in the processing parts of the complex spans 4. after removing multivariate outliers in all span tasks 5. after removing multivariate outliers in all span tasks and the processing parts of the complex spans 6. after removing outliers in depression (i.e., BDI-II score larger than 20) and stress (i.e., PSQ score larger or smaller than 3 SDs) 7. after removing outliers in depression (i.e., BDI-II score larger than 20) and stress (i.e., PSQ score larger or smaller than 3 SDs) as well as those participants reporting medicament intake in the last 24 hours or alcohol, drug, caffeine or nicotine intake in the last 2 hours before a session
SEM approach	1. latent-change model without constraints 2. latent-change model with the constraints of positive factor loadings 3. latent-change model with the constraints of positive error variances 4. latent-change model with the constraints of positive factor loadings and error variances 5. latent-change model without the covariance between the latent-change factors in case of bivariate latent-change model 6. latent-change model without the measures which were used to assess attentional control and which showed a significant synchrony effect at the individual-task level 7. second-order latent growth model

Note. Structural equation modeling was performed only if the sample size was larger than 80 participants. The procedure presented in the main text is displayed in bold. SD = standard deviation; BDI-II = Beck Depression Inventory II (Hautzinger et al., 2006); PSQ = Perceived Stress Questionnaire (Fliege et al., 2005); SEM = Structural Equation Modeling.

^a Participants with a score smaller than 50 were classified as evening types, whereas participants with a score larger than 50 were classified as morning types. For participants with a score of 50, they were classified as morning or evening types according to the question: "One hears about 'morning types' and 'evening types'. Which one of these types do you consider yourself to be?" The potential responses were: "Definitely a morning type", "Rather more a morning type than an evening type", "Rather more an evening type than a morning type", and "Definitely an evening type" (see Griefahn et al., 2001).

tive findings with regards to a synchrony effect on cognition is most likely coincidental. Therefore, the synchrony effect at the individual-task level is rather the exception than the norm.

On the latent-variable level, we were also not able to observe any evidence for a synchrony effect on cognition. Specifically, we used structural equation modeling to assess short-term maintenance and attentional control separately for on- and off-peak times, and we estimated the synchrony effect as a latent change between both peaks. Across the different analyses we run, we found no systematic synchrony effect for short-term maintenance or attentional control. The reasons were multiple: (1) The models did not provide good fit statistics or were not fully identified. (2) The factors did not represent shared variances across the measures. (3) The latent synchrony effect was not significant. There were only two cases with a specific combination of data transformation, participants' selection and trimming procedure where the model provided good fit statistics, substantial common variance, and a significant – but small – latent synchrony effect. However, consistent with the results at the individual-task level, these two cases

are the exception and represent most likely coincidental findings. Therefore, the systematic picture of the SEM approach unequivocally reveals no evidence for a general and robust synchrony effect on attentional control and short-term maintenance. Overall, the present results are mostly in line with the findings challenging the synchrony effect on cognition (e.g., Ceglarek et al., 2021; Heimola et al., 2021; Knight & Mather, 2013; Lewandowska et al., 2017; Li et al., 1998; Matchock & Mordkoff, 2008).

The results show no evidence for a general and robust the synchrony effect: Is this conclusion warranted?

One may argue that our conclusion is not warranted because (a) we did not have sufficient power in the present study, (b) there was an unbalanced number of morning and evening types, and (c) there were differences in the recruitment and the testing of the participants. We next discuss each of these concerns in detail.

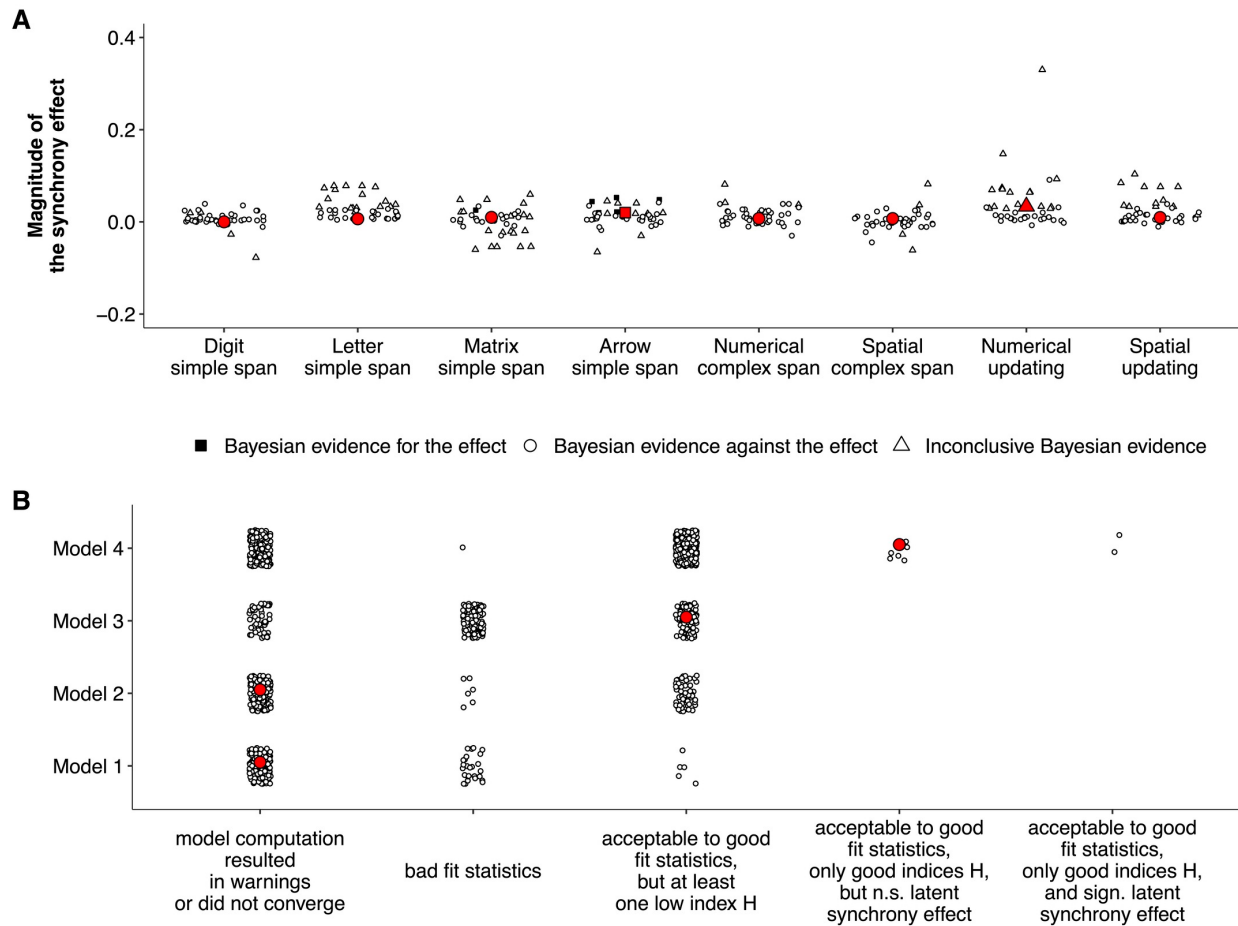


Figure 4. Synchrony Effect: Overview of the Results from the Multiverse-Analysis Approach

Note. Panel A: Magnitude of the synchrony effect across short-term and working-memory tasks. Panel B: Results from the latent-change models estimating the synchrony effect for short-term maintenance and attentional control. A model fit was considered as good if the Bentler's comparative fit index (CFI) was larger than .95 and the standardized root-mean-square residual (SRMR) was smaller than .08. A model fit was considered as acceptable if the CFI ranged from .90 to .95 and the SRMR was smaller than .08. Otherwise, the model fit was considered as bad. The index *H* was considered as low if it was smaller than .70. Otherwise, it was considered as good. For both panels, each datapoint represents the results combining the different data transformations, participants' selections, trimming procedures, and analyses listed in Table 9. The results reported in the text are presented in red. N.s. = not significant; sign. = significant.

Sufficient power?

Because we put forward that the small sample sizes typically used in previous research are an issue, a first concern may be whether in the present study, we had sufficient power to detect a true synchrony effect. This is particularly important because we focused on the subsample consisting of 191 moderate to definite chronotypes. The analyses on the complete sample including the 446 participants were reported in the multiverse-analysis approach.

For the analyses at the individual-task level, we can estimate the adequacy of the sample size by using the recommended sample sizes put forward by Brysbaert (2019). According to his recommendations, a sample of 70 participants should be sufficient to identify an effect in a within-subject design using a *t*-test with an effect size of .40, a power of .90, and an alpha level of .05. If a Bayesian approach is used, a sample of 131 participants should be sufficient to detect a synchrony effect with a Bayes Factor larger than 10 (see the Table 9 on p. 27, Brysbaert, 2019). Together, this indicates that with our sample of 191 partici-

pants, we had enough power to detect a synchrony effect in the present study.

The question is now whether we had enough power to have strong evidence against the synchrony effect. According to Brysbaert (2019), a sample of 1800 participants with moderate to definite chronotypes would have been necessary to find a Bayes factor larger than 10, thus indicating strong evidence against the synchrony effect. This is interesting for two reasons. First, it may explain why most Bayes factors in favor of the absence of the synchrony effect were larger than 3 but still smaller than 10 in the present study. Second, the requirement of a sample size consisting of 1800 participants with moderate to definite chronotypes puts forward the difficulty of having strong evidence against the synchrony effect. If we applied the same ratio between moderate to definite chronotypes and neutral types as the one observed in the present study (i.e., 43%), the goal of testing 1800 participants with moderate to definite chronotypes would require the recruitment and testing of more than 4200 participants. In the current research field, even if an online study is used, this seems difficult to implement.

Therefore, we must acknowledge that our sample size was not large enough to provide strong evidence against a synchrony effect. Nevertheless, the conditions to provide strong evidence seem currently unrealistic. Moreover, the present study includes one of the largest sample sizes used so far (cf. [Table 1](#)), and the results at the individual-task level showed positive evidence against a synchrony effect with a Bayes factor larger than 3 for most tasks. Together, this warrants our conclusion that the synchrony effect is not as robust and general as previously thought.

For the analyses at the latent-variable level, we applied the recent approach put forward by Bader et al. (2022) to determine the required sample size for bifactor models. We opted for this approach because it takes into account the complexity of the bifactor models. The details of the approach are presented in Appendix C. The results show that a sample size of 191 participants is sufficient for estimating all our models with a high rate of proper convergence. Moreover, accurate parameter estimations are expected for most models (in particular for Model 3 and Model 4 in which we observed good fit statistics). Together, this suggests that the power in the present study was sufficient in all analyses to warrant our conclusion.

Imbalance between morning and evening types: Is this an issue in the present study?

A second concern may be the imbalance between morning and evening types we reported in the present study. Our results showed that 131 young adults were categorized as morning types, whereas 60 young adults were categorized as evening types. One may wonder whether this unbalanced number of morning and evening types is an issue for the statistical analyses. This is not the case because we used a within-subject design. That is, because all participants were tested in the morning and evening, all participants were tested at both on-peak and off-peak times. Such a design allows us to collapse the within-subject variables Session (morning vs. evening) and Chronotype (morning type vs. evening type) into a single within-subject variable Testing time (on peak vs. off peak). Thus, for each task, we were able to compute a paired *t*-test with the within-subject variable “Testing time” with the two levels “on peak” and “off peak”. The same applied for the latent-change model, which can be conceptualized as a paired *t*-test to latent constructs. Therefore, the imbalance between morning and evening types does not affect the conclusions for all our analyses.

We nevertheless agree that it could seem unexpected to observe more young adults with morning types than with evening types. This contrasts with the typical finding reported in previous research according to which young adults have almost exclusively an evening chronotype (see, e.g., May et al., 1993; May & Hasher, 1998b). However, a close inspection of the normative data reported by Griefahn et al. (2001, see their first figure) showed rather a balanced number of morning and evening types for young adults. Moreover, we found a balanced number of morning and evening types in young adults in two datasets (see Rothen & Meier, 2016; and Rothen, 2023). In an unpublished

dataset (Rothen, 2015), we even observed more morning types than evening types in young adults, which is line with the present results. For the sake of transparency, we present in Appendix D figures displaying the number of morning, neutral, and evening types as a function of age for each of these datasets. Critically, this discrepancy in the findings can be explained. That is, in contrast to previous research in which American college students were tested, we tested in all our datasets – including the present one – young European adults coming from the general population. Thus, the previous result of more evening types in young adults might have been biased by testing a very specific population with very specific habits. This explanation is so far purely speculative, asking for further research in order to be tested directly and thoroughly.

Different recruitment and testing procedures?

A third concern may be whether we used different recruitment and testing procedures than those used in previous research. For example, a quick look at [Table 2](#) might suggest that the recruitment and testing did not work well because of errors made by the students who recruited and tested the participants. According to this table, 65 participants were not completing the experiment at the correct time, six participants did the tasks in the wrong order, and 69 participants did not perform all tasks. However, these issues cannot be solely attributed to students’ errors. For example, because the study was performed online and each participant started the experiment on its own, it is possible that some participants did not recognize the importance of being tested at the specific time, although it was carefully explained by the students. Similarly, most participants did the tasks in the wrong order because they typed a wrong participant’s number in one of the sessions. However, because the order of the tasks was determined by the participant’s number, this typo resulted in a different task order across the sessions. Thus, this issue was rather a result of a participant’s mistake or misunderstanding. Finally, participants were excluded because at least one task was missing. Here, there might be several reasons leading to a missing task. For example, a crash could have occurred, thus leading to the exclusion of the task. We also excluded a task if a break longer than 3 minutes was observed during the execution of a task. To our best knowledge, we are not aware of any online studies controlling for such fine-grained experimental settings. Together, all the exclusion criteria makes our online study more similar to laboratory settings. Moreover, in all tasks, we observed performance similar to previous research (e.g., in the mean, reliability and correlation estimates; cf. Kane et al., 2004; Rey-Mermet et al., 2019). This suggests that the quality of our data is similar to the quality observed in previous studies using in-lab testing. Therefore, the testing procedure cannot account for the discrepancy between our results and previous research.

Open questions and next steps

The present results showed no evidence for a general and robust synchrony effect in young adults for short-term

memory and working-memory tasks and for the constructs of short-term maintenance and attentional control. However, there are still some remaining questions. For example, previous research emphasizing the synchrony effect has frequently compared performance in young and older adults (e.g., Intons-Peterson et al., 1998, 1999; Knight & Mather, 2013; Li et al., 1998; May, 1999; May et al., 2005; May & Hasher, 1998b; Rowe et al., 2009). Therefore, the synchrony effect might be more general and robust in older adults. However, the results of the present study ask to be cautious when designing the next studies by opting for a design in which older adults are tested at both on- and off-peak times (see Rothen & Meier, 2016, for an example). Thus, the critical comparison testing the synchrony effect would be within the same group of older adults. In addition to increasing the statistical power, this has the advantage that the synchrony effect would not be affected by performance differences between young and older adults caused, for example, by the general slowing typically observed in older adults or by different speed-accuracy trade-offs (see, e.g., Salthouse, 1979, 1996).

Another open question concerns the impact of moderators on the synchrony effect, such as the sleep-wake history (see, e.g., Dijk & von Schantz, 2005). In the present study, we followed previous research and we did not explicitly control for prior sleep-wake history. The reason is that such control might have introduced a bias in the recruitment and testing of the participants so that the sample would no longer have been recruited and tested as described in the seminal studies. Therefore, whereas the present results indicate no robust and general synchrony effect for a population with no specific sleep-wake history, they do not preclude a synchrony effect in populations with disturbed sleep-wake history, such as shift-workers with off-peak working schedules. Therefore, future studies should be designed to determine to what extent prior sleep-wake history, in particular disturbed sleep-wake history, affects the synchrony effect.

The present results show no evidence for a robust and general synchrony effect when the chronotype was assessed using the Morningness-Eveningness Questionnaire (i.e., the D-MEQ, Griefahn et al., 2001). In the present study, we opted for such a method because we followed previous seminal studies (see, e.g., Intons-Peterson et al., 1998; May, 1999; May et al., 1993). However, these results do not preclude that circadian arousal can have an impact on cognition when other methods are used. For example, such an impact has been reported when participants were tested using a constant routine protocol or a forced desynchronization protocol (see Schmidt et al., 2007; Valdez, 2018, for reviews). In the constant routine protocol, the core body temperature, melatonin, cortisol, and cognitive performance are measured at regular intervals for a minimum of 24 hours, and participants are asked to stay awake with reduced motor activity. In a forced desynchronization protocol, participants are asked to adjust their sleep-wake cycle to, for example, 28h-period, and cognitive performance is measured at different times of this period. When these methods are used, the results indicate an impact of circa-

dian arousal on cognitive constructs, such as attention and working memory (see, e.g., Ramírez et al., 2006; Valdez et al., 2005). Therefore, these methods seem more appropriate to detect a true impact of circadian arousal on cognition.

Conclusion

More generally, the results of the present study convey an important message which applies to psychological science and scientific research more widely. Namely, reliable and robust insights into cognitive processes result from well-powered experimental designs in which several tasks measure the same underlying constructs. Had we conducted our study with only one of the two tasks which revealed positive findings, we would have come to the wrong conclusion that the synchrony effect does exist. Moreover, given our large sample, the within-subject design and the reliability of our measures, we would have interpreted the findings as very robust and representative. As our results show no evidence for a general and robust synchrony effect across the different tasks and the different levels of analyses (individual-task vs. latent-variable level), such an interpretation of our results is not warranted. Given the overall picture of our results, we must conclude that the synchrony effect is not as robust and general as previously thought.

Contributions

Contributed to conception and design: ARM, NR
 Contributed to acquisition of data: ARM
 Contributed to analysis of data: ARM
 Contributed to interpretation of data: ARM, NR
 Drafted and/or revised the article: ARM, NR
 Approved the submitted version for publication: ARM, NR

Acknowledgements

We thank Elisabeth Schoch, Stefanie Tangeten, and Daniel Fitze as well as the students who took part in the course M08 in fall term 2020 and in the course M1 in spring term 2021 for their help in data collection. We also thank Niels Kempkens for testing the reproducibility of the analyses.

Funding information

ARM is currently supported by a grant from the Swiss National Science Foundation (Grant 100014_207865). NR is currently supported by a grant from the Swiss National Science Foundation (Grant 10001CM_204314).

Competing interests

We have no known conflict interest to disclose.

Data accessibility statement

All deidentified data, experiment codes, research materials, analysis codes, and results are publicly accessible on the Open Science Framework (OSF) at <https://osf.io/ngfxv>. This study's design and its analysis were pre-registered on

OSF. The preregistration can be accessed at <https://osf.io/tywu7>.

Submitted: May 26, 2023 PDT, Accepted: August 29, 2023 PDT



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

References

- Allen, P. A., Grabbe, J., McCarthy, A., Bush, A. H., & Wallace, B. (2008). The Early Bird Does Not Get the Worm: Time-of-Day Effects on College Students' Basic Cognitive Processing. *The American Journal of Psychology*, 121(4), 551–564. <https://doi.org/10.2307/20445486>
- Aust, F., & Barth, M. (2020). *papaja: Prepare reproducible APA journal articles with R Markdown*. <https://github.com/crsh/papaja>
- Baddeley, A. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology*, 63(1), 1–29. <https://doi.org/10.1146/annurev-psych-120710-100422>
- Bader, M., Jobst, L. J., & Moshagen, M. (2022). Sample Size Requirements for Bifactor Models. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(5), 772–783. <https://doi.org/10.1080/10705511.2021.2019587>
- Bennett, C. L., Petros, T. V., Johnson, M., & Ferraro, F. R. (2008). Individual Differences in the Influence of Time of Day on Executive Functions. *The American Journal of Psychology*, 121(3), 349–361. <https://doi.org/10.2307/20445471>
- Bodenhausen, G. V. (1990). Stereotypes as Judgmental Heuristics: Evidence of Circadian Variations in Discrimination. *Psychological Science*, 1(5), 319–322. <https://doi.org/10.1111/j.1467-9280.1990.tb00226.x>
- Bonnefond, A., Gisselbrecht, D., Hoeft, A., Eschenlauer, R., Muzet, A., & Tassi, P. (2003). Cognitive performance in middle-aged adults as a function of time of day and task load. *Neurobiology of Sleep-Wakefulness Cycle*, 3, 1–8.
- Brain function of night owls and larks differ, study suggests.* (2019, February). BBC News. <https://www.bbc.com/news/health-47238070>
- Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, 2(1), 16. <https://doi.org/10.5334/joc.72>
- Ceglarek, A., Hubalewska-Mazgaj, M., Lewandowska, K., Sikora-Wachowicz, B., Marek, T., & Fafrowicz, M. (2021). Time-of-day effects on objective and subjective short-term memory task performance. *Chronobiology International*, 38(9), 1330–1343. <https://doi.org/10.1080/07420528.2021.1929279>
- Ceurstemon, S. (2020). *Introducing Chronotypes*. Owl + Lark. <https://owllark.com/journal/feature/introducing-chronotypes-part-2/>
- Chelminski, I., Petros, T. V., Plaud, J. J., & Ferraro, F. R. (2000). Psychometric properties of the reduced Horne and Ostberg questionnaire. *Personality and Individual Differences*, 29(3), 469–478. [https://doi.org/10.1016/S0191-8869\(99\)00208-1](https://doi.org/10.1016/S0191-8869(99)00208-1)
- Cohen, D. (2014, January). *Are you a lark or an owl?* BBC News. <https://www.bbc.com/news/health-25777978>
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5), 769–786. <https://doi.org/10.3758/bf03196772>
- Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, 1(1), 42–45. <https://doi.org/10.20982/tqmp.01.1.p042>
- Dijk, D.-J., & von Schantz, M. (2005). Timing and Consolidation of Human Sleep, Wakefulness, and Performance by a Symphony of Oscillators. *Journal of Biological Rhythms*, 20(4), 279–290. <https://doi.org/10.1177/0748730405278292>
- Draheim, C., Pak, R., Draheim, A. A., & Engle, R. W. (2022). The role of attention control in complex real-world tasks. *Psychonomic Bulletin & Review*, 29(4), 1143–1197. <https://doi.org/10.3758/s13423-021-02052-2>
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, 128(3), 309–331. <https://doi.org/10.1037/0096-3445.128.3.309>
- Fabbri, M., Mencarelli, C., Adan, A., & Natale, V. (2013). Time-of-day and circadian typology on memory retrieval. *Biological Rhythm Research*, 44(1), 125–142. <https://doi.org/10.1080/09291016.2012.656244>
- Fabbri, M., Natale, V., & Adan, A. (2008). Effect of time of day on arithmetic fact retrieval in a number-matching task. *Acta Psychologica*, 127(2), 485–490. <https://doi.org/10.1016/j.actpsy.2007.08.011>
- Fliege, H., Rose, M., Arck, P., Walter, O. B., Kocalevent, R.-D., Weber, C., & Klapp, B. F. (2005). The Perceived Stress Questionnaire (PSQ) Reconsidered: Validation and Reference Values From Different Clinical and Healthy Adult Samples. *Psychosomatic Medicine*, 67(1), 78–88. <https://doi.org/10.1097/01.psy.0000151491.80178.78>
- Ghisletta, P., & McArdle, J. J. (2012). Latent Curve Models and Latent Change Score Models Estimated in R. *Structural Equation Modeling: A Multidisciplinary Journal*, 19(4), 651–682. <https://doi.org/10.1080/10705511.2012.713275>
- Goldstein, D., Hahn, C. S., Hasher, L., Wiprzycka, U. J., & Zelazo, P. D. (2007). Time of day, intellectual performance, and behavioral problems in Morning versus Evening type adolescents: Is there a synchrony effect? *Personality and Individual Differences*, 42(3), 431–440. <https://doi.org/10.1016/j.paid.2006.07.008>
- Griefahn, B., Kunemund, C., Brode, P., & Mehnert, P. (2001). Zur Validität der deutschen Übersetzung des Morningness-Eveningness-Questionnaires von Horne und Ostberg. The Validity of a German Version of the Morningness-Eveningness-Questionnaire Developed by Horne and Ostberg. *Somnologie*, 5(2), 71–80. <https://doi.org/10.1046/j.1439-054x.2001.01149.x>

- Hahn, C., Cowell, J. M., Wiprzycka, U. J., Goldstein, D., Ralph, M., Hasher, L., & Zelazo, P. D. (2012). Circadian rhythms in executive function during the transition to adolescence: The effect of synchrony between chronotype and time of day. *Developmental Science*, 15(3), 408–416. <https://doi.org/10.1111/j.1467-7687.2012.01137.x>
- Hasher, L., Chung, C., May, C. P., & Foong, N. (2002). Age, time of testing, and proactive interference. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale*, 56(3), 200–207. <https://doi.org/10.1037/h0087397>
- Hautzinger, M., Keller, F., & Kühner, C. (2006). *Beck Depression Inventar II (BDI 2)*. Harcourt Test Service.
- Heimola, M., Paulanto, K., Alakuijala, A., Tuisku, K., Simola, P., Ämmälä, A.-J., Räisänen, P., Parkkola, K., & Paunio, T. (2021). Chronotype as self-regulation: Morning preference is associated with better working memory strategy independent of sleep. *SLEEP Advances*, 2(1), zpab016. <https://doi.org/10.1093/sleepadvances/zpab016>
- Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2019). *Lab.js: A free, open, online study builder*. <https://doi.org/10.31234/osf.io/fqr49>
- Horne, J. A., & Östberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology*, 4(2), 97–110. <https://doi.org/10.3109/07420529709001458>
- Horne, J. A., & Östberg, O. (1977). Individual differences in human circadian rhythms. *Biological Psychology*, 5(3), 179–190. [https://doi.org/10.1016/0301-0511\(77\)90001-1](https://doi.org/10.1016/0301-0511(77)90001-1)
- Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4), 424–453. <https://doi.org/10.1037/1082-989X.3.4.424>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519.909540118>
- Intons-Peterson, M. J., Rocchi, P., West, T., McLellan, K., & Hackney, A. (1998). Aging, optimal testing times, and negative priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(2), 362–376. <https://doi.org/10.1037/0278-7393.24.2.362>
- 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(1), 23–40. <https://doi.org/10.1037/0278-7393.25.1.23>
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2021). *semTools: Useful tools for structural equation modeling*. <https://CRAN.R-project.org/package=semTools>
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. *Journal of Experimental Psychology: General*, 133(2), 189–217. <https://doi.org/10.1037/0096-3445.133.2.189>
- Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin*, 144(11), 1147–1185. <https://doi.org/10.1037/bul0000160>
- Kievit, R. A., Brandmaier, A. M., Ziegler, G., van Harmelen, A.-L., de Mooij, S. M. M., Moutoussis, M., Goodyer, I. M., Bullmore, E., Jones, P. B., Fonagy, P., Lindenberger, U., & Dolan, R. J. (2018). Developmental cognitive neuroscience using latent change score models: A tutorial and applications. *Developmental Cognitive Neuroscience*, 33, 99–117. <https://doi.org/10.1016/j.dcn.2017.11.007>
- Knight, M., & Mather, M. (2013). Look Out—It’s Your Off-Peak Time of Day! Time of Day Matters More for Alerting than for Orienting or Executive Attention. *Experimental Aging Research*, 39(3), 305–321. <https://doi.org/10.1080/0361073x.2013.779197>
- Kühnberger, A., Fritz, A., & Scherndl, T. (2014). Publication Bias in Psychology: A Diagnosis Based on the Correlation between Effect Size and Sample Size. *PLOS ONE*, 9(9), e105825. <https://doi.org/10.1371/journal.pone.0105825>
- Lara, T., Madrid, J. A., & Correa, Á. (2014). The Vigilance Decrement in Executive Function Is Attenuated When Individual Chronotypes Perform at Their Optimal Time of Day. *PLOS ONE*, 9(2), e88820. <https://doi.org/10.1371/journal.pone.0088820>
- Lewandowska, K., Wachowicz, B., Marek, T., Oginska, H., & Fafrowicz, M. (2017). Would you say “yes” in the evening? Time-of-day effect on response bias in four types of working memory recognition tasks. *Chronobiology International*, 35(1), 80–89. <https://doi.org/10.1080/07420528.2017.1386666>
- Li, K. Z. H., Hasher, L., Jonas, D., Rahhal, T. A., & May, C. P. (1998). Distractibility, circadian arousal, and aging: A boundary condition? *Psychology and Aging*, 13(4), 574–583. <https://doi.org/10.1037/0882-7974.13.4.574>
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3), 519–530. <https://doi.org/10.1093/biomet/57.3.519>
- Matchock, R. L., & Mordkoff, T. J. (2008). Chronotype and time-of-day influences on the alerting, orienting, and executive components of attention. *Experimental Brain Research*, 192(2), 189–198. <https://doi.org/10.1007/s00221-008-1567-6>
- May, C. P. (1999). Synchrony effects in cognition: The costs and a benefit. *Psychonomic Bulletin & Review*, 6(1), 142–147. <https://doi.org/10.3758/bf03210822>
- May, C. P., & Hasher, L. (1998a). Synchrony effects in inhibitory control over thought and action. *Journal of Experimental Psychology: Human Perception and Performance*, 24(2), 363–379. <https://doi.org/10.1037/0096-1523.24.2.363>

- May, C. P., & Hasher, L. (1998b). Synchrony effects in inhibitory control over thought and action. *Journal of Experimental Psychology: Human Perception and Performance*, 24(2), 363–379. <https://doi.org/10.1037/0096-1523.24.2.363>
- May, C. P., Hasher, L., & Foong, N. (2005). Implicit memory, age, and time of day. *Psychological Science*, 16(2), 96–100. <https://doi.org/10.1111/j.0956-7976.2005.00788.x>
- May, C. P., Hasher, L., & Stoltzfus, E. R. (1993). Optimal Time of Day and the Magnitude of Age Differences in Memory. *Psychological Science*, 4(5), 326–330. <http://doi.org/10.1111/j.1467-9280.1993.tb00573.x>
- McDermott, L. M., & Ebmeier, K. P. (2009). A meta-analysis of depression severity and cognitive function. *Journal of Affective Disorders*, 119(1–3), 1–8. <https://doi.org/10.1016/j.jad.2009.04.022>
- Miyake, A., & Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Current Directions in Psychological Science*, 21(1), 8–14. <http://doi.org/10.1177/0963721411429458>
- Morey, R. D. (2008). Confidence Intervals from Normalized Data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, 4(2), 61–64. <https://doi.org/10.20982/tqmp.04.2.p061>
- Morey, R. D., & Rouder, J. N. (2021). *BayesFactor: Computation of bayes factors for common designs*. <http://CRAN.R-project.org/package=BayesFactor>
- Muthén, L. K., & Muthén, B. O. (2002). How to Use a Monte Carlo Study to Decide on Sample Size and Determine Power. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(4), 599–620. https://doi.org/10.1207/s15328007sem0904_8
- Nuijten, M. B., Wetzels, R., Matzke, D., Dolan, C. V., & Wagenmakers, E.-J. (2014). A default Bayesian hypothesis test for mediation. *Behavior Research Methods*, 47(1), 85–97. <https://doi.org/10.3758/s13428-014-0470-2>
- Parsons, S. (2020). *Splithalf; robust estimates of split half reliability*. <https://doi.org/10.6084/m9.figshare.5559175.v5>
- Parsons, S., Kruijt, A.-W., & Fox, E. (2019). Psychological Science Needs a Standard Practice of Reporting the Reliability of Cognitive-Behavioral Measurements. *Advances in Methods and Practices in Psychological Science*, 2(4), 378–395. <https://doi.org/10.1177/2515245919879695>
- Petros, T. V., Beckwith, B. E., & Anderson, M. (1990). Individual differences in the effects of time of day and passage difficulty on prose memory in adults. *British Journal of Psychology*, 81(1), 63–72. <https://doi.org/10.1111/j.2044-8295.1990.tb02346.x>
- Pink, D. H. (2018). *When: The Scientific Secrets of Perfect Timing*. Penguin Publishing Group.
- Pornprasertmanit, S., Miller, P., Schoemann, A., Jorgensen, T. D., & Quick, C. (2021). *Simsem: SIMulated Structural Equation Modeling*.
- R Core Team. (2021). *R: A language and environment for statistical computing* (3.6.3). <https://www.R-project.org/>
- Raftery, A. E. (1995). Bayesian Model Selection in Social Research. *Sociological Methodology*, 25, 111. <https://doi.org/10.2307/271063>
- Ramírez, C., Talamantes, J., García, A., Morales, M., Valdez, P., & Menna-Barreto, L. (2006). Circadian rhythms in phonological and visuospatial storage components of working memory. *Biological Rhythm Research*, 37(5), 433–441. <https://doi.org/10.1080/09291010600870404>
- Revelle, W. (2021). *Psych: Procedures for psychological, psychometric, and personality research*. Northwestern University. <https://CRAN.R-project.org/package=psych>
- Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(4), 501–526. <https://doi.org/10.1037/xlm0000450>
- Rey-Mermet, A., Gade, M., Souza, A. S., von Bastian, C. C., & Oberauer, K. (2019). Is executive control related to working memory capacity and fluid intelligence? *Journal of Experimental Psychology: General*, 148(8), 1335–1372. <https://doi.org/10.1037/xge0000593>
- Rey-Mermet, A., Singh, K. A., Gignac, G. E., Brydges, C. R., & Ecker, U. K. H. (2020). Interference control in working memory: Evidence for discriminant validity between removal and inhibition tasks. *PLOS ONE*, 15(12), e0243053. <https://doi.org/10.1371/journal.pone.0243053>
- Rey-Mermet, A., Singmann, H., & Oberauer, K. (2021). Neither measurement error nor speed-accuracy trade-offs explain the difficulty of establishing attentional control as a psychometric construct: Evidence from a latent-variable analysis using diffusion modeling. <https://doi.org/10.31234/osf.io/3h26y>
- Rock, P. L., Roiser, J. P., Riedel, W. J., & Blackwell, A. D. (2013). Cognitive impairment in depression: A systematic review and meta-analysis. *Psychological Medicine*, 44(10), 2029–2040. <https://doi.org/10.1017/S0033291713002535>
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, 21(2), 137–150. <https://doi.org/10.1037/met0000045>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Rothen, N. (2015). *Data collected during an undergraduate course on experimental practice* [Data set]. University of Bern.
- Rothen, N. (2023). *No evidence for differential effects of chronotype and time of day on controlled and automatic processes at memory retrieval* [Talk]. The 65th Conference of Experimental Psychologists, Trier, Germany.
- Rothen, N., & Meier, B. (2016). Time of day affects implicit memory for unattended stimuli. *Consciousness and Cognition*, 46, 1–6. <https://doi.org/10.1016/j.concog.2016.09.012>

- Rowe, G., Hasher, L., & Turcotte, J. (2009). Short article: Age and synchrony effects in visuospatial working memory. *Quarterly Journal of Experimental Psychology*, 62(10), 1873–1880. <https://doi.org/10.1080/17470210902834852>
- Saenger, J., Bechtold, L., Schoofs, D., Blaszkewicz, M., & Wascher, E. (2014). The influence of acute stress on attention mechanisms and its electrophysiological correlates. *Frontiers in Behavioral Neuroscience*, 8. <https://doi.org/10.3389/fnbeh.2014.00353>
- Salthouse, T. A. (1979). Adult age and the speed-accuracy trade-off. *Ergonomics*, 22(7), 811–821. <https://doi.org/10.1080/00140137908924659>
- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, 103(3), 403–428. <https://doi.org/10.1037/0033-295x.103.3.403>
- Savage, A. (2020, August). *Larks and Owls- What Do Those Chronotypes Tell You About Your Cognition & Mental Health?* BBC News. <https://www.linkedin.com/pulse/larks-owls-what-do-those-chronotypes-tell-you-your-cognition-savage>
- Schmidt, C., Collette, F., Cajochen, C., & Peigneux, P. (2007). A time to think: Circadian rhythms in human cognition. *Cognitive Neuropsychology*, 24(7), 755–789. <https://doi.org/10.1080/02643290701754158>
- Schmidt, C., Collette, F., Reichert, C. F., Maire, M., Vandewalle, G., Peigneux, P., & Cajochen, C. (2015). Pushing the Limits: Chronotype and Time of Day Modulate Working Memory-Dependent Cerebral Activity. *Frontiers in Neurology*, 6. <https://doi.org/10.3389/fneur.2015.00199>
- Schmidt, C., Peigneux, P., Leclercq, Y., Sterpenich, V., Vandewalle, G., Phillips, C., Berthomier, P., Berthomier, C., Tinguely, G., Gais, S., Schabus, M., Desseilles, M., Dang-Vu, T., Salmon, E., Degueldre, C., Balteau, E., Luxen, A., Cajochen, C., Maquet, P., & Collette, F. (2012). Circadian Preference Modulates the Neural Substrate of Conflict Processing across the Day. *PLoS ONE*, 7(1), e29658. <https://doi.org/10.1371/journal.pone.0029658>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. *Dialogue: The Official Newsletter of the Society for Personality and Social Psychology*, 26, 4–7.
- Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2021). *Afex: Analysis of factorial experiments*. <https://CRAN.R-project.org/package=afex>
- Soper, D. S. (2018). *A-priori Sample Size Calculator for Structural Equation Models [Software]*. <http://www.danieldsoper.com/statcalc>
- Starcke, K., Wiesen, C., Trozke, P., & Brand, M. (2016). Effects of Acute Laboratory Stress on Executive Functions. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.00461>
- Tabachnick, B. G., & Fidell, L. S. (2019). *Using Multivariate Statistics* (7th ed.). Pearson.
- Valdez, P. (2018). Homeostatic and circadian regulation of cognitive performance. *Biological Rhythm Research*, 50(1), 85–93. <https://doi.org/10.1080/09291016.2018.1491271>
- Valdez, P., Ramírez, C., García, A., Talamantes, J., Armijo, P., & Borrani, J. (2005). Circadian rhythms in components of attention. *Biological Rhythm Research*, 36(1–2), 57–65. <https://doi.org/10.1080/09291010400028633>
- Van Opstal, F., Aslanov, V., & Schnelzer, S. (2022). Mind-wandering in Larks and Owls: The Effects of Chronotype and Time of Day on the Frequency of Task-unrelated Thoughts. *Collabra: Psychology*, 8(1), 57536. <https://doi.org/10.1525/collabra.57536>
- von Bastian, C. C., Blais, C., Brewer, G. A., Gyurkovics, M., Hedge, C., Kałamała, P., Meier, M. E., Oberauer, K., Rey-Mermet, A., Rouder, J. N., Souza, A. S., Bartsch, L. M., Conway, A. R. A., Draheim, C., Engle, R. W., Friedman, N. P., Frischkorn, G. T., Gustavson, D. E., Koch, I., ... Wiemers, E. A. (2020). *Advancing the understanding of individual differences in attentional control: Theoretical, methodological, and analytical considerations*. <https://doi.org/10.31234/osf.io/x3b9k>
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14(5), 779–804. <https://doi.org/10.3758/bf03194105>
- West, R., Murphy, K. J., Armilio, M. L., Craik, F. I. M., & Stuss, D. T. (2002). Effects of Time of Day on Age Differences in Working Memory. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 57(1), P3–P10. <https://doi.org/10.1093/geronb/57.1.p3>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>
- Yang, L., Hasher, L., & Wilson, D. E. (2007). Synchrony effects in automatic and controlled retrieval. *Psychonomic Bulletin & Review*, 14(1), 51–56. <https://doi.org/10.3758/bf03194027>
- Yaremenko, S., Sauerland, M., & Hope, L. (2021). Eyewitness identification performance is not affected by time-of-day optimality. *Scientific Reports*, 11(1), 3462. <https://doi.org/10.1038/s41598-021-82628-z>
- Yoon, C. (1997). Age Differences in Consumers' Processing Strategies: An Investigation of Moderating Influences. *Journal of Consumer Research*, 24(3), 329–342. <https://doi.org/10.1086/209514>

Appendices

Appendix A. Deviations from the preregistration

In the present study, there are a few deviations from the preregistration. These are listed below.

Changes in the terminology

In the preregistration, we used the term “interference control”. However, to be more in line with recent research (von Bastian et al., 2020), we opted for the term “attentional control” in the present study.

Furthermore, in the preregistration, we used the term “extreme groups” regarding the chronotypes. This terminology may refer to the moderate to definite chronotypes or only the definite chronotypes. To avoid any confusion, we avoided “extreme groups” in the present study by referring either to moderate to definite chronotypes or to definite chronotypes.

Number of hypotheses for the construct “maintenance”

In the preregistration, we formulated only one hypothesis for the construct “maintenance” due the state of research at that time. Because more recent research has established a difficulty of finding a synchrony effect for short-term memory and working-memory tasks in the meantime (see, e.g., Ceglarek et al., 2021; Heimola et al., 2021), we formulated two contradictory hypotheses for this construct in the present study.

Recruitment in courses M08 and M1

In the preregistration, participants were planned to be recruited by the students from UniDistance Suisse who take part in the course M08. Because the sample was not completed after the course M08, participants were also recruited during the course M1.

Duration between sessions 2 and 3

In the preregistration, Sessions 2 and 3 were planned to be separated by maximally one week. During testing, this duration was extended if the participant could not perform the session as planned (e.g., because of illness).

Sample size

In the preregistration, the target sample size was 453 participants. After applying the inclusion criteria described in [Table 2](#), we were close to this target sample size because the sample consisted of 446 participants. Please note that in the present study, we mainly report the analyses on the subsample including the participants with moderate to definite morning and evening types ($N = 191$). The reason is that we aim to be consistent with the seminal research published by May and colleagues (see, e.g., Intons-Peterson et al., 1998; May, 1999; May et al., 1993). The analyses on

the full sample including all chronotypes ($N = 446$) are presented as part of the multiverse-analysis approach.

Statistical power

In the preregistration, the statistical power was described to be set to .95. This was a typo. A power of .90 was used to determine the sample size.

Dependent measures

In the preregistration, the dependent measures were error rates. For the sake of simplicity, we used accuracy rates ($= 1 - \text{error rates}$) in the present study.

Furthermore, in the preregistration, we used the term “partial-credit unit scoring method” to refer to the computation of the dependent measures as the proportion of memoranda recalled at the correct position. This was an error because our computation of the dependent measures corresponds to the partial-credit load scoring method. This was corrected in the present study.

Reliability of the measures

Reliability was planned to be calculated by adjusting split-half correlations with the Spearman–Brown prophecy formula. To have more stable estimates, we computed permutation-based split-half reliability estimates with the Spearman–Brown prophecy formula (Parsons et al., 2019).

Zero-order correlations

The upper and lower confidence intervals of the correlations were planned to be computed from a bootstrapping procedure with 10,000 random samples. We simplified the analyses by using the R-package *psych* (Version 2.2.9; Revelle, 2021) and by reporting the upper and lower confidence intervals computed with this package.

Furthermore, in the preregistration, the Bayes factors for the correlations were planned to be estimated using the BayesMed package (Nuijten et al., 2014) with default prior scales. Because this package could no longer be used, the Bayes factors were estimated using the R-package *BayesFactor* (Version 0.9.12.4.4; Morey & Rouder, 2021) with default prior scales.

Fit measures

In the preregistration, we forgot to mention that RMSEA values are less preferable when the sample size includes less than 250 participants (Hu & Bentler, 1998). The information was added in the present study.

Furthermore, in the preregistration, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) were introduced as inference criteria. The goal of these fit measures is to compare two models. Because our results did not allow us to perform such model comparisons, these fit measures were not introduced in the present study. For the same reason, we did not perform χ^2 difference ($\Delta\chi^2$) tests on nested models and the Bayesian hy-

pothesis test with the BIC approximation (Wagenmakers, 2007).

Factor reliability

In the preregistration, the factor reliability was planned to be assessed using a coefficient ω . Because this coefficient cannot be computed for all models we estimated, we preferred computing the index H . This index could be computed for all models.

Measurement models

In the preregistration, three measurement models – that is, Model 1, Model 2, and Model 3 – were introduced. These models were estimated. The results are available on OSF.

Second-order latent growth curve modeling approach

In the preregistration, analyses using second-order latent growth curve models were introduced as main analyses. Because these model assessments never resulted in acceptable or good fit statistics, we opted for another approach with the estimation of latent-change models.

Appendix B. Time-of-day effects

We investigated the time-of-day effect, that is, the difference in performance between the morning and evening sessions. Thus, irrespective of the chronotype, performance in the morning was compared to performance in the evening. Consistent with the analyses on the synchrony effect, the analyses were first performed at the individual-task level and then at the latent-variable level.

At the individual-task level, the descriptive results are displayed in [Figure B1](#), and the results from the null hypothesis significance testing (NHST) and Bayesian approach are described in [Table B1](#). For most tasks, no time-of-day effect was observed. Only in the digit simple span and in the numerical updating task, Bayesian evidence was inconclusive regarding the presence or the absence of the effect. However, the effect sizes were small for these tasks.

At the latent-variable level, the results are presented in [Figure B2](#). Consistent with the results on the synchrony effect, the results showed a difficulty of measuring maintenance and attentional control at the latent-variable level. When these were modeled satisfactorily (i.e., with acceptable to good fit statistics and good indices H), the time-of-day effect was in most cases not significant. When, however, it was significant, the latent change was small (maximum value for the unstandardized intercept of the latent-change factor = 0.02).

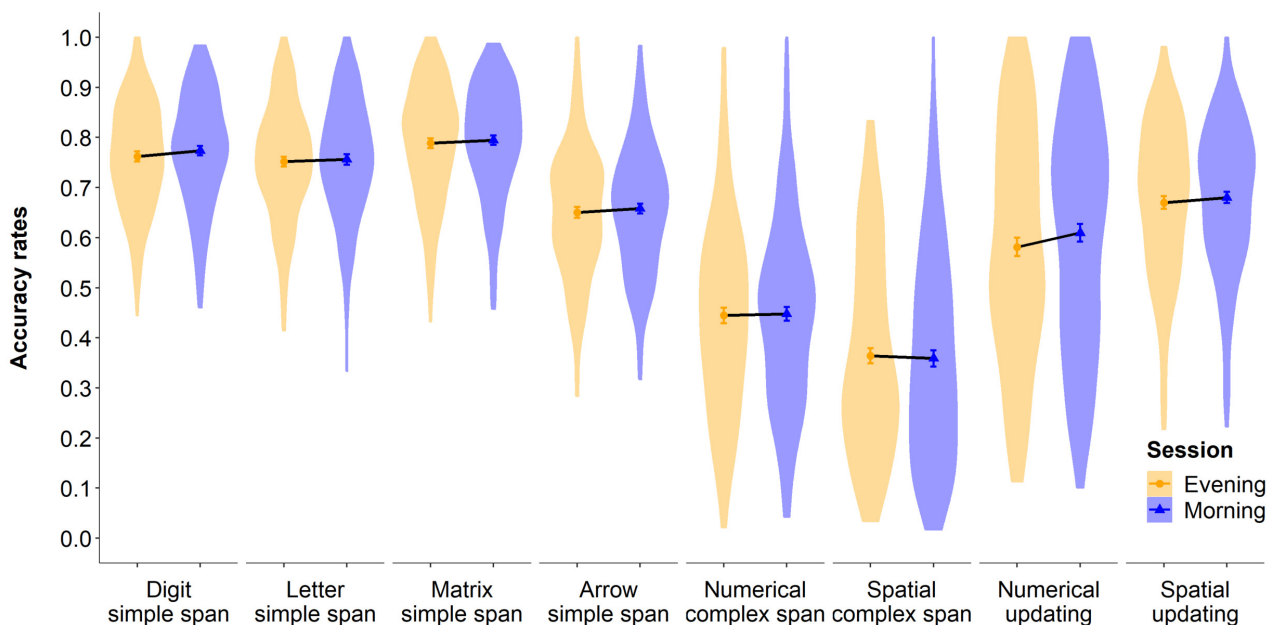


Figure B1. Time-Of-Day Effect (i.e., Mean Accuracy Difference between Morning and Evening Sessions) for Each Short-Term Memory and Working-Memory Task

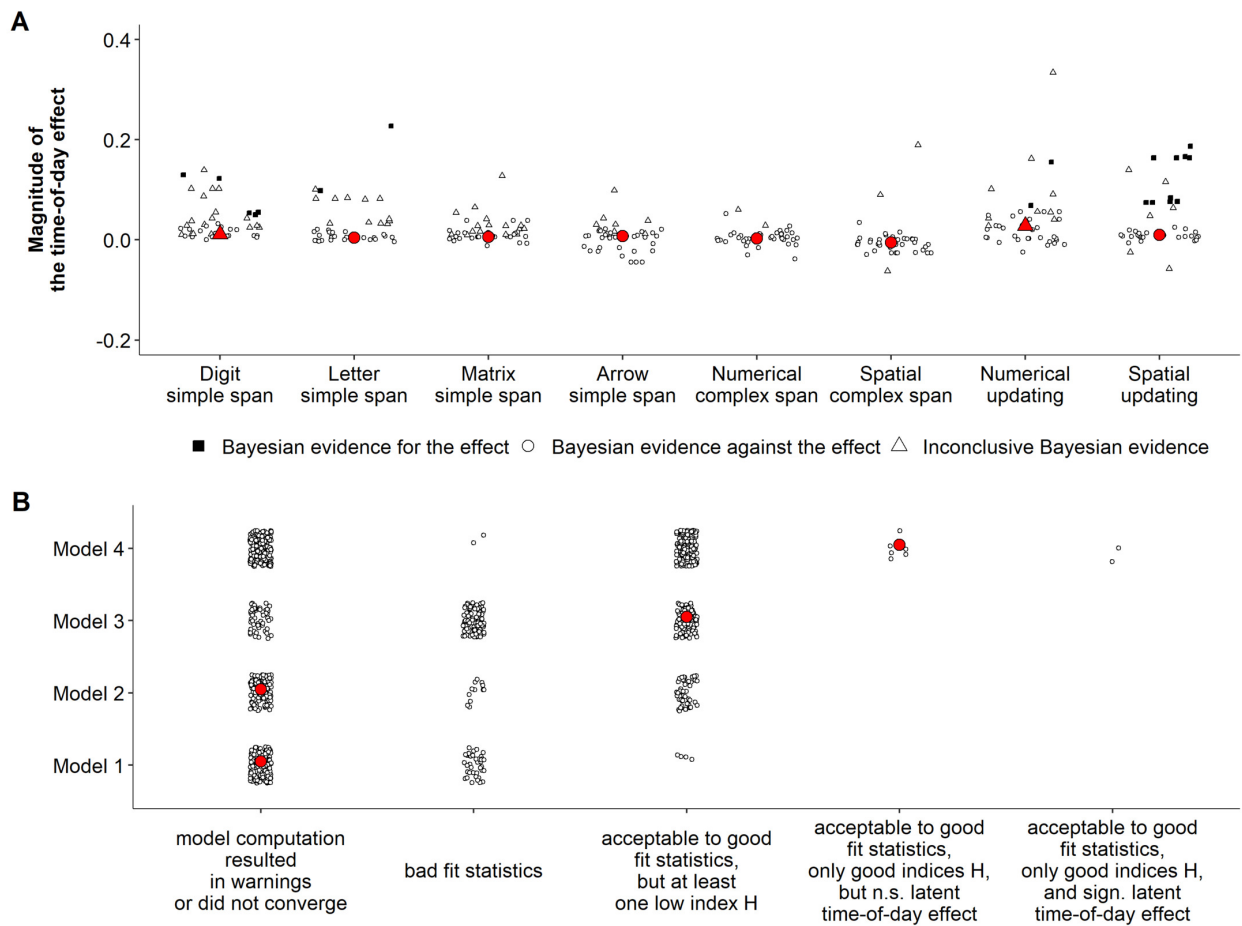
Note. Error bars represent within-subject confidence intervals (Cousineau, 2005; Morey, 2008). Shaded areas display data density plots.

Table B1. Time-Of-Day Effect: Inferential Statistical Values for the *t*-Tests Comparing Morning and Evening Sessions for Each Short-Term Memory and Working-Memory Task Separately, and Bayes Factors (BF) from the Bayesian *t*-Tests.

Task	<i>t</i> (190)	<i>p</i>	Cohen's <i>d</i>	BF ₁₀	BF ₀₁
Digit simple span	1.77	.079	0.13	0.37	2.70
Letter simple span	0.60	.546	0.04	0.10	10.34
Matrix simple span	0.90	.368	0.07	0.12	8.29
Arrow simple span	1.01	.313	0.07	0.13	7.49
Numerical complex span	0.27	.789	0.02	0.08	11.94
Spatial complex span	-0.48	.632	0.03	0.09	11.05
Numerical updating	2.00	.047*	0.14	0.57	1.76
Spatial updating	1.22	.226	0.09	0.17	5.99

Note. For the sake of clarity, results with a BF₁₀ larger than 3 are presented in bold, whereas results with a BF₀₁ larger than 3 are presented in italics.

**p* < .05.

**Figure B2. Time-Of-Day Effect: Overview of the Results from the Multiverse-Analysis Approach**

Note. Panel A: Magnitude of the time-of-day effect across short-term memory and working-memory tasks. Panel B: Results from the latent-change models estimating the time-of-day effect for short-term maintenance and attentional control. A model fit was considered as good if the Bentler's comparative fit index (CFI) was larger than .95 and the standardized root-mean-square residual (SRMR) was smaller than .08. A model fit was considered as acceptable if the CFI ranged from .90 to .95 and the SRMR was smaller than .08. Otherwise, the model fit was considered as bad. The index *H* was considered as low if it was smaller than .70. Otherwise, it was considered as good. For both panels, each datapoint represents the results combining the different data transformations, participants' selections, trimming procedures, and analyses listed in Table 9. The results from the procedure reported in the text are presented in red. N.s. = not significant; sign. = significant.

Appendix C. Sample size requirements

We estimated the adequacy of our sample sizes by applying the approach from Bader et al. (2022) using the R package “simsem” (Pornprasertmanit et al., 2021). For the sake of completeness, we applied this approach for the subsample of 191 participants with moderate to definite chronotypes as well as for the full sample of 446 participants with all chronotypes. We computed all four models (i.e., Model 1, Model 2, Model 3, and Model 4). The factor loadings used to determine the target sample sizes were estimated from the loadings reported by Kane et al. (2004, see Figure 6 on p. 206). We used these factor loadings because the design used by Kane et al. (2004) was similar to the present study. That is, in both studies, a sample of young adults were asked to perform several short-term memory and working-memory tasks, and a partial-credit scoring procedure was used to compute the dependent variable of each task. According to Kane et al. (2004), one critical feature that supports the view that attentional control was extracted from the working memory tasks in their model was that the magnitude of factor loadings differ, depending on the type of task (short-term memory vs. working memory) and the type of factor (attentional-control factor vs. maintenance factors). This means that working-memory measures had higher factor loadings for the attentional-control factor than for the short-term memory measures. Conversely, the short-term memory measures had higher factor loadings for the maintenance factors than the working-memory measures. To reflect this feature in the selection of the factor loadings for our sample-size computations, we computed the median factor loadings separately for each task type (working memory vs. short-term memory) and each factor type (attentional control vs. maintenance). Accordingly, we selected .5, .8, .7, and .3 as factor loadings from the short-term memory measures and the working-memory measures to the attentional-control factor and maintenance factors, respectively.

Bader et al. (2022) put forward various criteria to evaluate whether the sample size is acceptable. First, the rate of proper convergence should be larger than 90%. Second,

the coverage – that is, the proportion of solutions in which 95% confidence interval covered the true population value – should be close to the .95. Third, the relative bias in the factor loadings and their estimated standard errors should be taken into account. That is, the absolute values of these parameters are considered as negligible if they are smaller than .05. They are considered as moderate if they range between .05 and .10. However, they are considered as strongly biased and thus unacceptable if they are larger than .10. Finally, the explained common variance (ECV) is computed in order to assess the proportion of common variance in the dependent variables accounted for by the general factor compared to the specific factors. Here, the goal is to obtain accurate ECV estimates by having average and relative ECV biases as small as possible (e.g., biases in absolute values smaller than .03).

The results are summarized in Table C1. As presented in this table, the different criteria were fulfilled for nearly all models with both sample sizes, thus indicating an acceptable sample size. There is only one exception. For Model 2 with the subsample of 191 participants, the rate of proper convergence was larger than 90%, thus indicating an acceptable convergence rate. However, the minimum coverage slightly deviated from the recommended value of .95 in comparison to the other computations. Moreover, the estimated standard errors of the factor loadings were strongly biased. In particular, they were underestimated (-0.24), suggesting that significant effects may be overestimated (Muthén & Muthén, 2002). However, the estimation of this model with our data reveals that the covariance matrix was not positive definite, thus indicating no proper convergence. Thus, the issue in this model is not in the accurate estimation of the parameters and their standard errors but in the proper convergence of the model. Together, this suggests that our sample sizes were sufficient to estimate all models with a high rate of convergence and to obtain accurate parameter estimates in Models 1, 3 and 4 (see Model 2 for an exception).

Table C1. Measures for the Different Criteria used to Assess the Adequacy of the Sample Size for all Models (i.e., Model 1, Model 2, Model 3, and Model 4) and for the Subsample Including Only Moderate to Definite Morning and Evening Chronotypes ($N = 191$) and for the Full Sample Including All Chronotypes ($N = 446$)

Model	Sample size	Convergence rate	Coverage range	Relative bias range	Relative SE bias range	Average ECV bias	Relative ECV bias
Model 1	191	100.00	0.93 - 0.96	0 - 0.01	0 - 0.06	0.001	0.006
Model 2	191	96.35	0.91 - 0.96	0 - 0.02	0.01 - 0.24	-0.003	-0.013
Model 3	191	100.00	0.93 - 0.96	0 - 0.02	0 - 0.07	0.001	0.003
Model 4	191	99.80	0.93 - 0.96	0 - 0.02	0 - 0.06	0	0.001
Model 1	446	100.00	0.93 - 0.97	0 - 0.02	0 - 0.06	0	0.001
Model 2	446	100.00	0.93 - 0.96	0 - 0.01	0 - 0.07	-0.002	-0.006
Model 3	446	100.00	0.93 - 0.96	0 - 0.01	0 - 0.08	0	0.001
Model 4	446	100.00	0.93 - 0.96	0 - 0.01	0 - 0.1	0	0.001

Note. The ranges for the relative bias in the factor loadings and their standard errors are given in absolute values. SE = standard error; ECV = Explained common variance.

Appendix D. Chronoscore as a function of age

In our lab, we have collected two datasets in which we observed a balanced number of morning and evening types in young adults (see Rothen, 2023; Rothen & Meier, 2016). We have also collected two other datasets in which we observed the finding of more morning types than evening types in young adults (i.e., Rothen, 2015, and the dataset reported in the present study). The first three datasets were collected in the same way. That is, each dataset was collected in the context of a research methods class at the University of Bern. Second-year psychology students were asked to recruit and to test 16 participants each. All participants completed an online version of the Morningness-Eveningness Questionnaire (D-MEQ, Griefahn et al., 2001). Then, they were assigned to the experimental condition either in the morning (between 6:00 and 10:00) or in the evening (between 17:00 and 21:00). For the last dataset (i.e., the dataset from the present study), there were only a few modifications. First, the complete experiment was performed online. Second, students from UniDistance Suisse were asked to recruit eight participants each. Third, the testing times ranged from 07:30 to 10:00 for the morning session and from 16:30 to 19:00 for the evening session.

Scores from the D-MEQ are presented as a function of age in [Figures D1 to D4](#) for the four datasets, respectively. As shown in these figures, the correlation between D-MEQ scores and age was negligible for each dataset. Moreover, Bayes Factors (BFs) suggest positive evidence against the correlation in each case. Together, these results challenge the view that young adults have almost exclusively an evening chronotype (see, e.g., May et al., 1993; May & Hasher, 1998b)

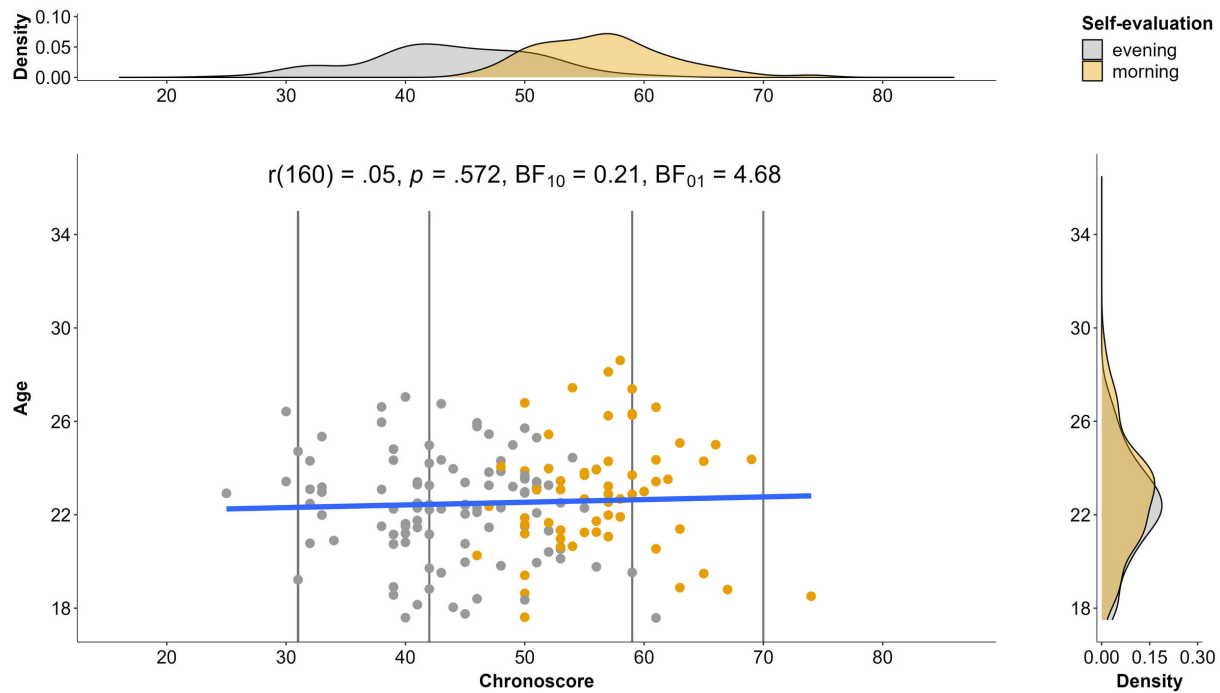


Figure D1. Dataset from Rothen and Meier (2016) – Chronoscore as a Function of Age

Note. The central panel displays the relationship between self-declared eveningness/morningness, age and the achieved score in the German version of the Morningness-Eveningness Questionnaire (D-MEQ; Griefahn et al., 2001). Self-declared eveningness/morningness is based on the last question of the D-MEQ: 'One hears about 'morning types' and 'evening types'. Which one of these types do you consider yourself to be?' The potential answers were: 'Definitely a morning type', 'Rather more a morning type than an evening type', 'Rather more an evening type than a morning type', 'Definitely an evening type'. Thus, participants who classified themselves in this last question as evening types are shown in grey and participants who classified themselves as morning types are shown in orange. Vertical lines indicate the thresholds for the definite evening, moderate evening, neutral, moderate morning, and definite morning range (from left to right). These thresholds were based on the standard chronotype classification (Chelminski et al., 2000; Griefahn et al., 2001). A random jitter of ± 0.5 was added to each datapoint in order to help visual inspection of the data. The blue line indicates the correlation between age and D-MEQ score. The top panel depicts the density of the distribution of the D-MEQ scores, and the left panel depicts the density of the distribution of the participants' age. In this dataset, we observed 22 morning types (with a D-MEQ score ranging from 59 to 86), 100 neutral types (with a D-MEQ score ranging from 42 to 58) and 38 evening types (with a D-MEQ score ranging from 16 to 41). Self-declaration according to the last question in the D-MEQ revealed 61 morning types (orange dots) and 99 evening types (grey dots). This figure was adapted from the figure 2 from Rothen and Meier (2016).

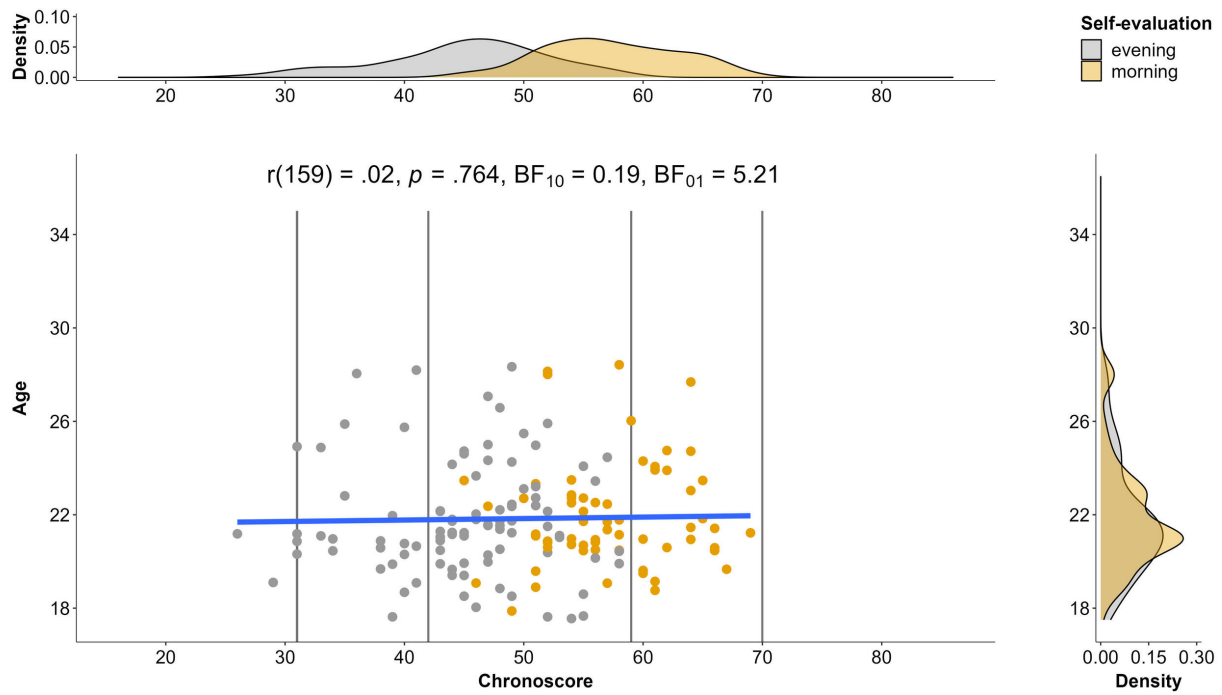


Figure D2. Dataset from Rothen (2023) – Chronoscore as a Function of Age

Note. See [Figure D1](#) note for details. In this dataset, we observed 26 morning types (with a D-MEQ score ranging from 59 to 86), 108 neutral types (with a D-MEQ score ranging from 42 to 58) and 26 evening types (with a D-MEQ score ranging from 16 to 41). Self-declaration according to the last question in the D-MEQ revealed 66 morning types (orange dots) and 94 evening types (grey dots).

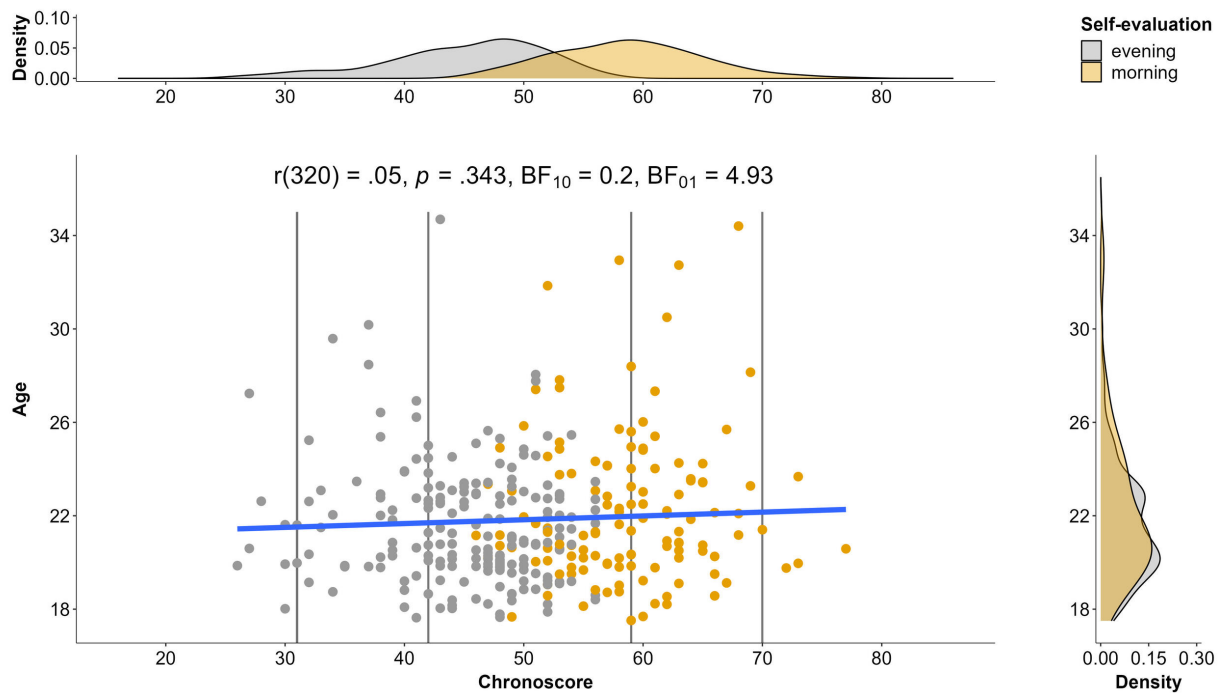


Figure D3. Unpublished Dataset from Rothen (2015) – Chronoscore as a Function of Age

Note. See [Figure D1](#) note for details. In this dataset, we observed 64 morning types (with a D-MEQ score ranging from 59 to 86), 207 neutral types (with a D-MEQ score ranging from 42 to 58) and 49 evening types (with a D-MEQ score ranging from 16 to 41). Self-declaration according to the last question in the D-MEQ revealed 125 morning types (orange dots) and 195 evening types (grey dots).

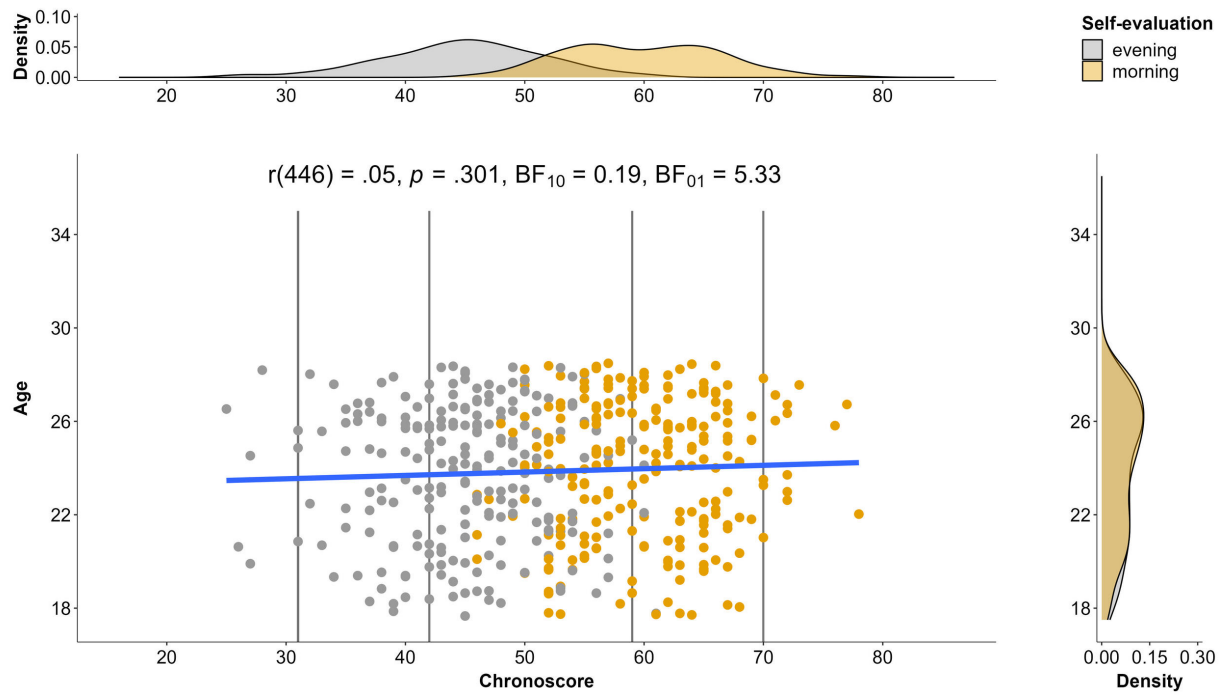


Figure D4. Dataset from the present study – Chronoscore as a Function of Age

Note. See [Figure D1](#) note for details. In this dataset, we observed 131 morning types (with a D-MEQ score ranging from 59 to 86), 255 neutral types (with a D-MEQ score ranging from 42 to 58) and 60 evening types (with a D-MEQ score ranging from 16 to 41). Self-declaration according to the last question in the D-MEQ revealed 231 morning types (orange dots) and 215 evening types (grey dots).

Supplementary Materials

Peer Review History

Download: https://collabra.scholasticahq.com/article/88337-the-interplay-of-time-of-day-and-chronotype-results-in-no-general-and-robust-cognitive-boost/attachment/182158.docx?auth_token=E3W-VcnYpP76zE5D_hYo

Cover Letter

Download: https://collabra.scholasticahq.com/article/88337-the-interplay-of-time-of-day-and-chronotype-results-in-no-general-and-robust-cognitive-boost/attachment/184752.pdf?auth_token=E3W-VcnYpP76zE5D_hYo

Cover Response Letter

Download: https://collabra.scholasticahq.com/article/88337-the-interplay-of-time-of-day-and-chronotype-results-in-no-general-and-robust-cognitive-boost/attachment/184753.pdf?auth_token=E3W-VcnYpP76zE5D_hYo
