Cognition 206 (2021) 104482

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Validity of attention self-reports in younger and older adults☆

Andra Arnicane[a,⁎], Klaus Oberauer[a], Alessandra S. Souza[a,b]

[a] University of Zurich, Switzerland
[b] University of Porto, Portugal

ARTICLE INFO

Keywords:
Mind-wandering
Aging
Attention appraisal
Self-report validity
Attention monitoring
Current concerns

ABSTRACT

Human attention is subject to fluctuations. Mind-wandering (MW) – attending to thoughts unrelated to the current task demands – is considered a ubiquitous experience. According to the *Control Failure x Concerns* view (McVay & Kane, 2010), MW is curbed by executive control, and task-irrelevant thoughts enter consciousness due to attentional control lapses. The generation of off-task thoughts is assumed to increase with higher number of personal concerns. Challenging this view, older adults report less MW than younger adults. Here, we addressed the hypothesis that older adults report less MW due to a lower ability to notice attention lapses and to appraise their current on-task focus. In an age-comparative study (N = 40 younger and N = 44 older adults) using a battery of three tasks spanning working memory, reading comprehension, and sustained attention, we assessed the correlation between the degree of self-reported on-task focus and task performance on a trial-by-trial basis. Younger and older adults' degree of on-task attention measured through thought probes was correlated equally strongly with performance across trials in all tasks, indicating preserved ability to monitor attentional fluctuations in healthy aging. Self-reported current concerns' number and importance did not differ across age, and they did not predict self-reported attention across tasks. Our study shows that lower rates of MW in aging do not reflect lower validity of older adults' attentional appraisal or lower levels of current concerns.

1. Introduction

Mind-wandering (MW) is defined as engaging in thoughts unrelated to the current demands of the external environment (Schooler et al., 2011). For example, whilst reading this introduction, the reader's attention may drift towards their upcoming vacation or be captured by a concept of personal interest mentioned in the text, leading them to embark on a new train of thought. Because these cognitions are unrelated to the task at hand (namely, reading and comprehending the text), they are considered as an instance of MW. MW is usually assessed via self-report using either experimenter-scheduled thought probes (referred to as probe-caught MW), participant-initiated self-report (referred to as self-caught MW), or with a combination of both (Schooler et al., 2011). MW is a frequent human experience; earlier research reported ca. 50% rates of MW (Killingsworth & Gilbert, 2010; but see Seli et al. (2018) for an alternative view).

The aim of the present research is to address a puzzle in the MW literature: According to one prominent theory, the *Control Failure x Concerns* view (McVay & Kane, 2009, 2012a), MW is a failure of attention control. Older adults are often assumed to have impaired attention control relative to young adults (Braver & West, 2008; Hasher & Zacks, 1988). Yet, older adults report less MW than younger adults (Frank, Nara, Zavagnin, Touron, & Kane, 2015; Jackson & Balota, 2012; Jordão, Ferreira-Santos, Pinho, & St. Jacques, 2019; Krawietz, Tamplin, & Radvansky, 2012; Maillet et al., 2018; McVay, Meier, Touron, & Kane, 2013; Seli, Maillet, Smilek, Oakman, & Schacter, 2017; Zavagnin, Borella, & De Beni, 2014). This raises the concern whether older adults' MW reports are valid due to possible age-related decline in metacognitive ability (Einstein & Mcdaniel, 1997; Jackson & Balota, 2012; McVay et al., 2013; Zavagnin et al., 2014). Accordingly, the main goal of the present study was to investigate the validity of older compared to younger adults' MW reports. We next introduce the *Control Failure x Concerns* view (McVay & Kane, 2009, 2012a) in more detail, and review the literature on mind-wandering and age differences therein. Then, we will introduce our research questions and experimental design.

1.1. MW as control failure

According to the *Control Failure x Concerns* view, MW is a failure of executive control to uphold the task goal (McVay & Kane, 2009, 2012a).

☆ Andra Arnicane, Klaus Oberauer, and Alessandra S. Souza, Department of Psychology, University of Zurich, Switzerland. This research was supported by a grant from the Velux Foundation to A. S. Souza (project n° 1053). K. Oberauer was supported by the University Research Priority Program "Dynamics of healthy aging" at the University of Zurich. We thank Gian-Luca Gubler and Vanessa Vallesi for assistance in data collection and coding.
⁎ Corresponding author at: Department of Psychology, University of Zurich, Switzerland.
E-mail address: a.arnicane@gmail.com (A. Arnicane).

https://doi.org/10.1016/j.cognition.2020.104482
Received 26 March 2020; Received in revised form 30 September 2020; Accepted 1 October 2020

Executive control (also known as attention control or executive attention) is a set of cognitive functions responsible for supervising and controlling thoughts and actions in the service of achieving current goals (Logan, 1985). In the *Control Failure x Concerns* view, executive control upholds the current task goal, thereby preventing task-irrelevant thoughts from entering consciousness (McVay & Kane, 2010). Task-irrelevant thoughts are assumed to be generated continuously outside of conscious awareness, and to enter consciousness when executive control lapses (McVay & Kane, 2010). Thought generation may increase, for instance, when environmental cues activate representations of personally relevant goals, namely *current concerns* (Klinger, 1971, 2009). The extent to which the environment (e.g., the lab setting of a psychology study) can trigger the person's current concerns may influence the amount of thoughts that are generated and that press forward to enter consciousness, thereby increasing MW rates (McVay et al., 2013).

If MW is a byproduct of failing to control attention, then its occurrence should correlate negatively with performance in cognitive tasks that require focused attention on the task goal, and on goal-relevant information. Support for this assumption comes from inter-individual correlation studies with younger adults, using both laboratory-based and experience sampling methods to assess MW frequency (Kane et al., 2007, 2016; McVay & Kane, 2009; McVay & Kane, 2012b, McVay & Kane, 2012a; Robison & Unsworth, 2015, 2017, 2018; Unsworth & McMillan, 2013, Unsworth & McMillan, 2014a, 2014b; Unsworth & Robison, 2017). Of note, some authors found the opposite relationship between MW and working memory capacity in younger adults when MW was probed in tasks with low attention demands (Levinson, Smallwood, & Davidson, 2012; Rummel & Boywitt, 2014). This finding in itself, however, does not challenge the view of MW as control failure, because with a low demanding task, MW may reflect a voluntary disengagement from the task that does not produce performance costs.

The link between MW and executive control can also be investigated on the intra-individual level by assessing the covariance between fluctuations in attentional reports and task-performance over time. Task performance is lower during time intervals when people report MW compared to when they report being task-focused. This has been observed in a variety of tasks, for example, reading comprehension (Frank et al., 2015; Jackson & Balota, 2012; Krawietz et al., 2012), the Sustained Attention to Response Task (SART) (Kane et al., 2007; McVay et al., 2013; McVay & Kane, 2012a), and visual working memory (Adam & Vogel, 2017; Krimsky, Forster, Llabre, & Jha, 2017; Unsworth & Robison, 2016).

1.2. MW and current concerns

The second component of McVay and Kane's (2010) account of MW are *current concerns*. The concerns represent the person's current life goals, and concern-related thoughts arise when cued by the environment (Klinger, 1971, 2009). Specifically, current concerns are defined as *"the state of an organism between two points in time: the point at which the organism becomes committed to pursuing a particular goal and the point at which it either consummates or disengages from the goal"* (Klinger, Barta, & Maxeiner, 1980, p.1223). According to this definition, engagement in an experimental task is also a person's current goal, which coexists with any other goals a person may entertain. Importantly, the *Control Failure x Concerns* view argues that resulting MW is an interaction of control failure and the extent to which the environment cues a persons' current concerns, proposing that older adults' reduced MW is a consequence of laboratory setting not overlapping with concerns of this age group (McVay et al., 2013, pp. 145, 146). There is some experimental evidence that priming personal concerns moderately increases peoples' MW rates. For example, McVay and Kane (2013) collected keywords of participants' current concerns in a pre-study session and embedded these in a SART with word stimuli. When probed after personal goal-related keywords, participants' MW rates were 4% higher than for probes succeeding the presentation of other peoples' keywords, suggesting that current concerns play a role in mind-wandering. However, a large experience sampling study found that also in the context of daily life older adults report less MW than the younger group (Maillet et al., 2018). Therefore, it is unlikely that the lab environment failing to cue older adults' concerns is responsible for their reduced MW rates. One possibility is that older adults have fewer concerns/goals than their younger counterparts, as older age is associated with fewer normative life events – biological, psychological and social life transitions that are highly correlated with chronological age (Cavanaugh & Blanchard-Fields, 2018), as well as with preference for different goals and higher selectivity of these (Carstensen, 1993, 1995). For example, Parks, Klinger, and Perlmutter (1989) administered a concerns questionnaire and found that older adults' mean number of concerns was lower than that of younger adults'. To our knowledge, no further MW studies to date have quantified current concerns across age groups. Accordingly, our second goal was to survey the current concerns of our participants to assess for potential age differences and their relationship with MW propensity.

1.3. Executive control and MW in aging

Executive control is assumed to decline with aging (Braver & West, 2008; Hasher & Zacks, 1988). Recent meta-analyses have shown that, instead of a uniform decline, some facets of executive control are impaired in the course of aging – for example, the ability to divide attention (Verhaeghen, 2011), whereas other facets, such as some varieties of inhibition, are preserved (Rey-Mermet & Gade, 2018). Working memory capacity – interpreted in the *Control Failure x Concerns* view as an indicator of executive control – is severely reduced in older compared to younger adults (Brockmole & Logie, 2013), leading to the prediction that older adults should experience more MW than younger adults. However, studies show that older adults consistently report less MW than younger adults both in laboratory-based studies (Frank et al., 2015; Jackson & Balota, 2012; Krawietz et al., 2012; McVay et al., 2013; Seli et al., 2017; Zavagnin et al., 2014) as well as in experience sampling in daily life (Maillet et al., 2018). On the assumption that executive control is compromised in older age, the findings of less MW reports in older adults challenge the *Control Failure x Concerns* view (McVay et al., 2013).

Given the inconsistency between MW reports and executive control performance in older adults, researchers have started to question whether older adults are able to notice and report instances of MW as well as younger adults (Einstein & Mcdaniel, 1997; Jackson & Balota, 2012;McVay et al., 2013; Zavagnin et al., 2014). The ability to monitor one's own thoughts is investigated in the field of meta-cognition. Studies in this field indicate that people are sometimes unable to introspect on their cognitive processes (Nisbett & Wilson, 1977) or mental experience (Schooler, 2002), which may constrain the validity of self-reports of MW. Furthermore, this ability may also vary over time within the same individual. For example, Seli, Jonker, Cheyne, Cortes, and Smilek (2015) addressed the question of MW report validity in a sample of younger adults. Participants responded to dichotomous attention probes (asking whether the person was "on-task" vs. MW), as well as indicated how confident they were in their reports' accuracy. Most of the time, participants expressed high confidence in their attention reports, and high-confidence attention reports predicted reaction time variability in the main task. But when accompanied by low confidence ratings, attention reports were not correlated with objective performance. These findings demonstrate that, at least in younger adults, the ability to appraise one's attention varies from time to time, and reductions in this ability yield reductions in the predictive value of MW reports (Seli et al., 2015). Therefore, if older adults have an impaired ability to monitor and report their attentional states, that could explain why their MW reports appear to dissociate from their task performance.

One investigation of the validity of older adults' MW self-reports has been conducted with eye-tracking (Frank et al., 2015). The authors assessed the relation between objective measures (reading comprehension score and oculomotor behavior) and self-reported attention focus in younger and older adults. Attention probes had eight response options, reflecting the categories of on-task focus, task-related interference (e.g., worry about one's performance on the task), and MW. Compared to younger adults, older adults reported less MW but higher rates of task-related interference, a pattern that replicates previous studies (McVay et al., 2013; Zavagnin et al., 2014). Overall, older adults reported more on-task focus than younger adults. Critically, on-task reports predicted higher comprehension test scores in both age groups, implying equal validity of MW reports across age. However, interactions of thought report category and age group were found for several eye-tracking measures. Specifically, three eye-tracking measures indicated disrupted processing when older adults reported MW. For younger adults, by contrast, the eye-tracking measures of disruptions in processing went along with task-related interference rather than MW reports. Frank et al. (2015) proposed that younger and older adults may have different thresholds for classifying thoughts as MW or interference. This assumption poses a challenge for the use of categorical thought report format in MW studies across age. It is possible that older adults report less MW simply because they use a more conservative criterion for declaring an attentional state as MW.

1.4. Concerns about the measurement of MW

A recent concern in the literature pertains to the diversity of operationalizing the MW construct, expressing concern for the comparability of results (Seli et al., 2018), as well as the heterogeneity in the types of MW probes across studies. These concerns motivated efforts to evaluate the validity of MW self-reports (Frank et al., 2015; Schubert, Frischkorn, & Rummel, 2019; Weinstein, 2017; Weinstein, De Lima, & van der Zee, 2017; Wiemers & Redick, 2019). For example, Weinstein et al. (2017) found that probe question wording emphasizing mind-wandering produced higher reported MW rates than when the probes inquired whether the participants were on-task. However, in another study, the manipulation of probe framing yielded comparable rates of reported MW (Schubert et al., 2019).

A further concern in the MW literature is that attention can be on task to a different degree at a given time. Seli et al. (2018) presented two groups of participants either a dichotomous probe (i.e., "on-task" vs "off-task"), or a five-point Likert scale with the following options: "(1) completely on-task, (2) mostly on-task, (3) both focused on-task and on something else, (4) mostly focused on something other than the task, and (5) completely focused on something other than the task". The dichotomous condition yielded an average MW rate of 40%. The 5-point-scale condition yielded varying MW rates, depending on which rating levels were coded as on-task (e.g., whether only "fully on-task" or also "mostly on-task" were included in the "task-focus" category). Importantly, MW rates in the 5-point rating analyses were always lower than in the dichotomous condition, except when only the "fully on-task" responses were contrasted with the four remaining response options. Thus, categorical and especially dichotomous assessment might result in inflated MW rate estimates (Mrazek et al., 2012; Seli, et al., 2018). This research highlights the issue that attention focus may often be a mix, or a rapid succession, of on- and off-task thoughts, and hence more adequately reported on a continuous scale. Additionally, the continuous-scale response format circumvents the possibility of age differences in the interpretation of, or attitudes towards, verbally defined thought categories of "mind-wandering" or "on-task". If the lower rate of reported MW in older adults arises from a more conservative criterion for classifying an attentional state as "mind-wandering", then using a continuous response scale removing the need for such a criterion should abolish that phenomenon: Older adults should no longer report stronger on-task focus than younger adults. Accordingly, one goal of the present study was to further evaluate the validity of MW reports in aging using a continuous attention scale.

1.5. The present study

Our main aim was to assess the relationship between attention reports and objective task performance measures in younger and older adults. At several times during three cognitive tasks, participants reported their degree of attentiveness to the task on a continuous rating scale ranging from one (completely off-task) to nine (completely on-task). This response format has been used with younger adults and is appropriate for registering small attention fluctuations that have a measurable effect on performance (Adam & Vogel, 2017; Unsworth & McMillan, 2014a). We chose this rating scale for two reasons. First, categorical, dichotomous report modes may pose difficulties for the assessment of MW report validity across age due to possible differences between age groups in how to classify a thought as MW (see Frank et al., 2015). Second, the continuous attention scale allows us to detect small fluctuations in attentional state which may go undetected in categorical response modes but result in performance disruptions on a trial-by-trial basis. A disadvantage of this scale is its inability to assess the contents of off-task cognitions. Some studies have indicated that some off-task contents, such as task-related interference, are higher in older adults compared to younger adults (Frank et al., 2015; McVay et al., 2013). These studies show, nevertheless, that inclusion of the task-related-interference option does not change the fact that older adults report more on-task thoughts than younger adults. Therefore, we believe that ignoring the nuances between different types of off-task reports does not compromise our ability to answer our main question, namely whether older adults are able to detect and report fluctuations on their attentional state.

We assessed the relationship between attention reports and performance in three tasks, implementing two widely used tasks in MW research, namely a reading task and the SART (Robertson, Manly, Andrade, Baddeley, & Yiend, 1997). Additionally, we tested participants on a full-report visual working memory task, which showed high sensitivity to in-the-moment attention fluctuations on performance accuracy (Adam & Vogel, 2017). Lastly, we collected participants' current concerns by the means of a paper-and-pencil survey, using a modified version of the questionnaire from Parks et al. (1989) study, and tested for age differences in concerns, and an effect of the amount of concerns on mean reported attention. We were interested to (a) replicate the age difference in the number of current concerns reported by Parks et al. (1989), and (b) explore the relationship of concerns and self-reported attention.

Our predictions were as follows. First, we expect to replicate the within-person covariation between fluctuations in task performance and fluctuations in attention self-reports previously observed in younger adults (Adam & Vogel, 2017; Krimsky et al., 2017). Specifically, we expect that, as younger adults' self-reported attention increases, task errors will decrease as illustrated in all panels of Fig. 1. Second, our main interest was to assess how older adults' self-reports compare to those of younger adults'.

A first possibility is that the relationship of older adults' self-reports with their performance can be described with the same intercept and slope as in younger adults, leading to an overlap in both age groups' regression lines, as illustrated in Fig. 1A. Alternatively, the intercept of older adults' reports could be shifted upwards relative to younger adults', whereas their task accuracy isn't. Then, at the same level of rated attention, older adults would commit more errors than their younger counterparts (Fig. 1B). Nevertheless, in the scenario in Fig. 1B attention reports are equally valid in both age groups, as the ratings covary with within-person fluctuation of performance. Lastly, if older adults' reports are not valid, their attentional reports will be uncorrelated (or less correlated) with performance, as depicted in Fig. 1C, resulting in an interaction between age and attention rating in predicting performance.

2. Methods

2.1. Sample size and participants

Our sample size was planned to match previous studies assessing attention report validity (Frank et al., 2015,) $N = 36$ younger, $N = 40$ older adults) and current concerns (Parks et al., 1989,) $N = 42$ younger, $N = 42$ older adults) across age.

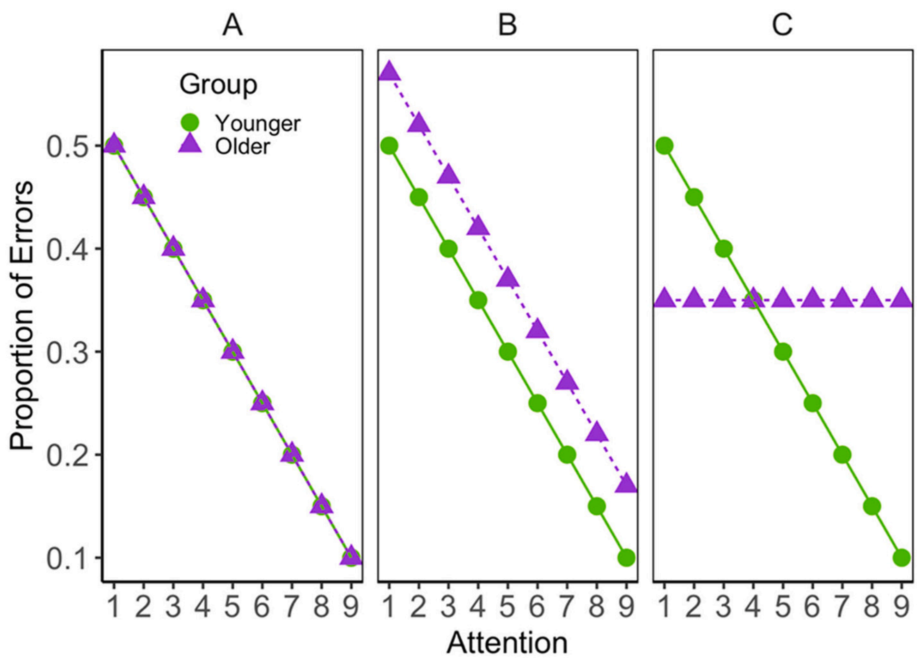

Fig. 1. Predictions for the relationship between proportion of errors and attention ratings for younger and older adults. Panel A illustrates the prediction under the hypothesis that attention reports are valid and the relationship between reports and performance is described by the same intercept and slope for both age groups. Panel B depicts the prediction under the hypothesis that older adults' attention reports are shifted upwards relative to the younger adults', but are valid in the sense that they covary with performance. Panel C illustrates the prediction under the hypothesis that older adults' reports do not covary with performance.

We tested 40 younger adults between 20 and 33 years (M_{Age} = 24.6, *SD* = 3.6, 33 women) and 44 community-dwelling older adults between 62 and 79 years (M_{Age} = 69.8, *SD* = 3.9, 22 women). All participants reported normal color vision and normal or corrected-to-normal visual acuity, and Swiss German or German as their mother tongue. Younger adults were students at a Swiss university or held a diploma comparable to a Swiss high-school certificate (Matura). There was no minimum education level requirement for older adults. Older adults completed the Mini-Mental State Examination (Folstein, Robins, & Helzer, 1983) and all obtained a score > 26. Two older adults elected not to attempt the visual working memory task because of difficulty. All collected data were included in the analyses.

Participants were compensated with either 15 Swiss Francs per hour or course credit (students only). Written informed consent was obtained from all participants prior to the start of the experiment, and participants were debriefed regarding the purpose of the study at the end. The study protocol is in line with the ethical guidelines of the institutional review board as established by the completion of a self-assessment checklist prior to the start of the study (checklist can be found here https://www.phil.uzh.ch/de/forschung/ethik.html).

2.2. Apparatus and stimuli

All tasks were programmed in MATLAB (The MathWorks, Natick, MA) using the Psychophysics Toolbox 3 (Brainard, 1997; Pelli, 1997). All tasks presented attention probes at quasi-randomized intervals, according to a schedule equal for all participants. An attention probe consisted of a question: “How attentive were you just now?” atop of a slider scale. On the left, the scale was anchored at: “1 = Not at all on task”, and on the right, at “9 = Fully on task”. Below the scale, an integer digit (from 1 to 9) was shown as soon as participants moved the mouse on the scale and changed according to the position of the mouse. Fig. 2 visualizes the attention probes (panel E) and the flows of events in the tasks (panels A-D).

2.2.1. Visual working memory task

Participants memorized the colors of a set of squares and at the end of a brief interval, they reported the color of each square by selecting it from a color checkerboard (Adam & Vogel, 2017). The memoranda were presented against a uniform grey background (RGB 128 128 128). The memory squares (side = 90 pixels) appeared in a subset of locations selected from an invisible 6 × 6 grid centered in the middle of the screen.

The colors of the memoranda were sampled without replacement from nine values (RGB): white (255 255 255), black (0 0 0), blue (0 0 255), cyan (0 255 255), green (0 255 0), yellow (255 255 0), 255 25 255 orange (255 128 0), red (255 0 0), and magenta (255 0 255). The item constellations were created with the following constraints. First, the memoranda locations within the array were separated by at least one grid cell. Second, each memory array in the experiment differed from all other arrays in at least two items (i.e., color-location assignments).

A trial started with a white fixation cross (size: 40 pixels) in the center of the screen (see Fig. 2A for timing of events). Next, positions of the items for the following trial were outlined in dark grey (RGB 112 112 112) against the grey background (placeholders). Then, the memoranda were presented simultaneously. Younger adults viewed five colors. In an attempt to equate the task difficulty across age groups, older adults were presented four colors[1]. During retention, the placeholders remained onscreen. At recall, a 3 × 3 color checkerboard (checkerboard side = 90 pixels), consisting of the nine colors listed above, appeared at the place of the to-be-recalled item. All items were tested sequentially in random order. To respond, participants clicked on one of the colors on the checkerboard. Already chosen colors were not available to be clicked on again in the same trial. Participants completed six practice trials with three attention probes, followed by 300 test trials with 60 attention probes. At test, the probes were spaced between three to seven trials apart. During practice, participants received performance feedback after each trial, and at test, after every 50 trials. Younger adults completed the task in approximately 60 min, and older adults in approximately 90 min. Supplementary analyses showed that longer task duration did not disadvantage older adults' attention ratings (see Appendix C).

2.2.2. Reading task

This task was modeled after Zedelius, Broadway, and Schooler

[1] This approach was taken to control for task difficulty across age groups. Capacity of WM declines across adult age (Brockmole & Logie, 2013). A recent study used an individual calibration procedure to equate memory load between younger and older adults in a delayed-estimation visual WM task (Loaiza & Souza, 2019). Their calibration results showed that older adults reached the same level of performance at a memory load of one item less than younger adults did. Based on these results, we presented older adults with one item less than the younger group, to preclude the confound of unequal task difficulty interacting with subjective attention ratings.

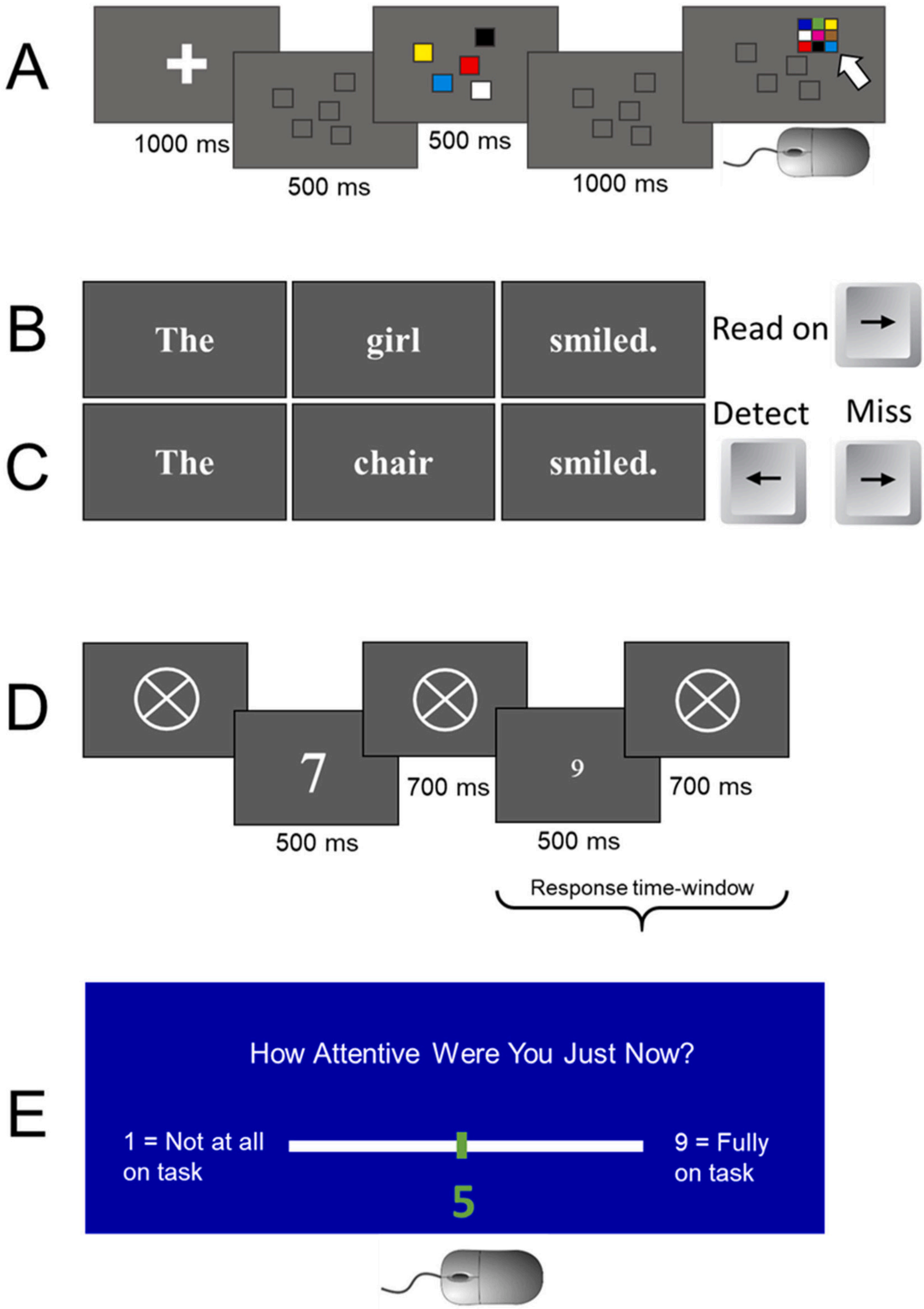

Fig. 2. The task event flows illustrate a trial in the visual working memory task (A), a sentence in the reading task with sensible (B) and gibberish (C) text, and trials in the SART (D). Panel E depicts the attention probe. The stimuli are not to scale.

(2015). Participants read a text, presented one word at a time in the screen-center (white against a grey 128 128 128 background in Times font, size = 34). Participants pressed the right arrow-key to display the next word. To prevent accidental skipping, minimum display time was 100 ms per word. Rereading previous text was not possible. Participants were instructed to read the text attentively and to report the presence of gibberish (nonsensical sentences) by pressing the left arrow key. After instruction, a 10-sentence example text with three attention probes followed. At test, participants read an abridged version of the introduction and the first five chapters from the book “Sommersprossen auf den Knien” by Maria Parr in German (length = 2781 words), rated by the publisher as appropriate for an age of 8–10 years.

The task phase presented 32 attention probes at quasi-random intervals (16 following normal text passages and 16 following gibberish passages). The text was interspersed with 16 gibberish episodes, each five sentences long, at quasi-random intervals. Gibberish was created by replacing story words with unrelated words. For example, the sentence "It is already a while ago" would become "It is pasture a try ago". No pseudo-words were used. The presented gibberish passages were previously rated in a pilot test ($N = 8$, younger adults) as gibberish with maximum confidence on a scale of one to four. Each gibberish passage was presented either until detection or until the full five sentences were read (whatever occurred first), then followed directly by an attention probe. Thereafter, the text was resumed to the normal version from where gibberish had started. The normal text was augmented with 16 added sentences presenting unique information (i.e., that occurred only once in the story) that matched the story framework (such as "The postman's name is Erik"), each followed by an attention probe.

After the end of the reading task, a long-term memory (LTM) test assessed memory for each of the 16 unique sentences by presenting four answer options: the correct sentence and three foils (e.g., "The postman's name is Erik / Bernd / Anno / Bengt"). Participants were instructed to click on the sentence they remembered reading in the story. Younger adults completed the task in approximately 30 min, and older adults in approximately 60 min. As in visual working memory task, supplementary analyses showed no disadvantage for older adults' attention ratings due to the longer task duration (see Appendix C).

2.2.3. SART

Participants were instructed to press the space bar for all digits between one and nine, except the digit three (aka no-go signal) (Robertson et al., 1997). The digits were presented in white against a grey background, in "Symbol" font, and in varying sizes (48, 72, 94, and 120 pts), according to the original task design by Robertson et al. (1997). All digits occurred with equal frequency. The digits remained onscreen for 500 ms and were masked by a disc (diameter = 60 pixels) with a cross for 700 ms. Responses were accepted until the onset of the next stimulus. Responses in the presence of the no-go stimulus were counted as commission errors. After every quarter of the trials, participants received feedback on their cumulative accuracy in percentage. In case of no response for 10 trials in a row, the task would pause with an alert message. Instructions refrained from emphasizing either speed or accuracy (Seli, Cheyne, & Smilek, 2012). Participants completed 18 example trials and 54 practice trials with three attention probes each. The test phase (900 trials) was interspersed with 20 attention probes. The digits, digit size, and attention ratings were presented according to a predetermined pseudo-random schedule, equal for all participants. Twenty blocks were created, comprising the digits one to nine between three to seven times. Thus, a block length of three was made of three concatenated vectors of one to nine, randomly shuffled within the block. Half of the blocks ended with the no-go digit. An attention probe followed after the end of each block. The task pace was automated and hence the task took approximately 40 min for both age groups.

At the end of each task, participants rated task difficulty and interest on a scale from one to nine. Thereafter, participants received feedback on their whole task performance.

2.2.4. Current concerns questionnaire

The Concerns Questionnaire was adapted with author's permission from the Personal Aspirations and Concerns Inventory (PACI; Cox, Klinger, & Fadardi, J. S. (Eds.)., 2011). The modifications were as follows: two life areas were removed for the reasons of privacy (*Intimacy and Sexual Matters* and *Religion and Spiritual Matters*), and area *Education and Training* was renamed *Education and Self-Development*, to address both age groups in equal measure. Furthermore, we introduced area *Fitness and Sport* to account for the high interest towards physical exercise in Switzerland, as well as area *Other* to accommodate concerns that did not fit into the provided categories. Participants were instructed to name concerns, if present, per area and rate each concern's importance on a scale from 0 (= barely important) to 10 (= highly important). Differently to the PACI, 10 instead of three rows per area were provided, and ratings instead of rankings of concerns were requested (see Appendix A for the instruction, list of life areas, and an example question). Participants were encouraged to write down "as many or as few concerns as they felt they had".

2.3. Procedure

Participants completed the tasks across two sessions. Before the start of each session, they received instructions regarding the attention probes and practiced placing attention ratings three times in the presence of the research assistant. The first session consisted of the visual working memory task, followed by the completion of the

Concerns Questionnaire. If they wished for, participants could take the questionnaire home and bring it back in the next session. The second session consisted of the reading task, a short break, and lastly the SART.

3. Results

3.1. Data analysis

Main inferential analyses were implemented as Bayesian hierarchical logistic regressions using the rstanarm package (Goodrich, Gabry, Ali, & Brilleman, 2018) in R (R Core Team, 2018). We set weakly informative Cauchy priors with location 0 and scale 5 (Gelman et al., 2013) for the intercept and the coefficients. The rstanam package internally adjusts the user-provided prior scales to account for the scales of the predictors (Goodrich et al., 2018). Hence, we report automatically adjusted prior scales when applicable. As a model quality statistic, we report the ratio of effective to total Markov Chain Monte Carlo (MCMC) samples. The effective sample size is the number of independent (non-autocorrelated) draws from the posterior distribution. Highly autocorrelated draws hinder the exploration of the parameter space by covering a small area of the posterior distribution and thus lead to unreliable estimates. The effective/total sample size ratio should be > 0.1 (Goodrich et al., 2018).

The analyses of task performance used the logit link and four chains of the MCMC sampling with 2000 iterations per chain. The first 1000 samples of each chain were discarded as warm-up, resulting in 4000 posterior distribution samples for each analysis. For all reported models, the $\widehat{R}$ statistic was < 1.02, indicating that the chains converged to the same posterior distribution. For inference, we assessed whether the 95% highest density interval (HDI) of the posterior parameter distributions included zero.

Furthermore, we also used Bayesian *t*-tests and Bayesian analyses of variance (BANOVAS) from the BayesFactor package (R. D. Morey, Rouder, Jamil, Forner, & Ly, 2018). We regarded Bayes factors (BF) between 3 and 10 as substantial, and larger than 30 as strong evidence in favor of H_1. Likewise, BFs between 0.3 and 0.10 were interpreted as substantial evidence for H_0 and BFs lower than 0.03 were regarded as strong evidence for H_0 (Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011). Ambiguous evidence for H_0 was indicated by BFs in the range 0.3 to 1 and for H_1 in the range 1 to 3.

All materials, data, and analysis scripts are available at https://osf.io/z7fdk/?view_only=9ca5956a57c144ea9947e39d088c3245.

3.2. Visual working memory task

Dependent measures in this task were (1) number of correctly reported items, (2) proportion of trials with lapses, and (3) attention ratings. Lapses reflect trials in which, presumably due to disengagement from the task, participants failed to store any information about the array, forcing them to guess (Adam & Vogel, 2017). Lapse trials are operationalized as trials with fewer than two correct responses (a single correct response could plausibly arise from guessing). Descriptive statistics of these measures are presented in Table 1 and histograms of response frequencies of the number of correctly reported memory items and attention ratings are presented in Appendix E (Fig. E1).

Table 1
Means and Standard Deviations (in Parentheses) of the Dependent Variables in the Visual Working Memory and Reading tasks.

	Visual working memory task				Reading task					
	Probed trials		Non-probed trials		Gibberish detection			LTM test		
Variable	Younger	Older	Younger	Older	Variable	Younger	Older	Variable	Younger	Older
# Correct Items	3 (1.28)	2.15 (1.06)	2.97 (1.26)	2.18 (1.08)	Detection Delay	1.23 (0.53)	1.12 (0.40)	Memory Accuracy	0.75 (0.43)	0.65 (0.47)
Proportion Lapses	0.13 (0.34)	0.25 (0.43)	0.13 (0.34)	0.24 (0.42)	Proportion Lapses	0.19 (0.39)	0.10 (0.30)	–	–	–
AR	6.20 (2)	6.93 (1.80)	–	–	AR	7.34 (1.44)	8.26 (0.91)	AR*	7.71 (1.25)	8.31 (0.91)
AR Reliability	0.82	0.88	–	–	AR Reliability	0.82	0.86	AR Reliability*	0.78	0.90

Note. AR – Attention Rating. Reliability estimates were calculated with the Spearman-Brown's correction. Visual working memory task: lapse ≤ 2 correct responses. Reading Task: detection delay is in averaged sentences. Lapse = detection later than in the first sentence. *Attention ratings listed for the LTM test were collected during the reading phase of non-gibberish passages.

First, we assessed whether task performance in probed trials differed from that in non-probed trials. A 2 (age group) × 2 (probed, non-probed) BANOVA on number of correctly reported items showed that performance in probed and non-probed trials did not differ (evidence for the probe factor was BF = 0.16). Evidence for an interaction did not reach the credible range (BF ratio = 2.92). The same analysis on lapse rates demonstrated that older adults lapsed more than younger adults (BF ratio = 20), but there was no evidence that this differed between probed and non-probed trials (evidence for inclusion of probe factor: BF ratio = 0.38 and for the interaction = 0.36). Together, the analyses suggested that the probed trial subset is an appropriate representation of the whole dataset.

Second, we assessed the relation between attention ratings and task performance. Fig. 3 displays scatterplots of the relation of attention ratings and the number of correctly recalled items and lapse proportion at the individual level. For number of correct items, increases in attention ratings were associated with increases in the number of correct recalls for both age groups. For lapse proportion, the converse was true: Increases in attention ratings were associated with lower lapse rates.

We modeled the relationship between attention and number of correct recalls in probed trials with a binomial Bayesian hierarchical regression model having attention rating and age group as (mean-centered) predictors. The adjusted prior scale for the coefficients was = [2.57, 5.00, 4.04], effective/total sample size ratio > 0.24, and the number of effective samples > 883. Means and HDIs of the posterior distributions are summarized in Table 2, and the model predictions are presented in Fig. 3 (dotted lines). The analysis showed two credible main effects: (1) higher reported attention predicted higher memory accuracy, and (2) older adults reported fewer correct items than younger adults did. The interaction term was not credible, indicating that the relationship of attention and accuracy was comparable in both age groups. The results did not change when mean number of correct items in the two or three trials before the probe was entered as dependent variable (Table 2).

Next, we analyzed whether attention ratings and age group predicted the probability of having lapsed in the trial just before. The adjusted prior scale for the coefficients was = [2.58, 5.00, 4.04], effective/total sample size ratio > 0.32, and the number of effective samples > 895. Paralleling the previous analysis, credible effects indicated that higher attention was associated with lower lapse rates, and that older adults experienced more lapses than younger adults (see Table 2). Again, the interaction term was not credible, indicating that the attention-performance association did not differ between age groups.

Effect sizes (standardized coefficients) for these models were computed with the "effectsize" package in R (Ben-Shachar, Makowski, & Lüdecke, 2020) and are reported in Table 2.

3.3. Reading task

After excluding key presses longer than four seconds (6% of the dataset), the average reading speed per word was $M_{Younger}$ = 328 ms (*SD* = 317) and M_{Older} = 579 ms (*SD* = 432). The main dependent variables in this task were (a) number of gibberish sentences read until detection, (b) long term memory (LTM) accuracy scores, and (c) attention ratings, visualized in Fig. 4. Table 1 presents descriptive statistics for these variables.

Only two out of 1344 gibberish occurrences went undetected. Most occurrences were detected in the first sentence (85.19%). Given that most of the gibberish was detected early, we reduced the number of levels of our predicted variable by re-coding the data such that gibberish detection during the second sentence or later was defined as a lapse. Fig. 4 displays gibberish detection lapses and proportion of correct responses in the LTM test as a function of attention ratings. A Bayesian *t*-test showed that older adults had fewer gibberish detection lapses than younger adults, BF_{10} = 15.

First, we analyzed gibberish detection lapses as a function of attention rating and age group with a binomial Bayesian hierarchical regression (adjusted prior scale for the coefficients = [3.90, 5.00, 3.28], effective/total sample size ratio > 0.10, number of effective samples > 368). Higher attention ratings credibly predicted fewer gibberish detection lapses (see parameter's HDI in Table 2). The HDI of the age group predictor and the interaction term both included zero, indicating that, firstly, the model did not identify a credible main effect of age group on lapse rate, and secondly, that attention ratings were similarly predictive for both groups' task performance.

Because the *t*-test had supported age group difference in gibberish detection but the main analysis did not, we hypothesized that the group effect was fully accounted for by the higher attention ratings of older adults. We repeated the analysis with only age group as a predictor, and the group term was now credible (mean = −0.61; 95% HDI = [−0.90, −0.32]). In other words, when attention ratings were not included as a predictor, the analysis revealed that older adults had fewer detection lapses than younger adults, but this effect was accounted for by including attention ratings as predictors.

Second, we analyzed accuracy in the LTM test. Binary response accuracy for each story fact presented directly before an attention probe in the non-gibberish text was entered in a binomial Bayesian hierarchical regression with attention rating and age group as mean-centered predictors (adjusted prior scale = [4.45,5.00,3.13], effective/total sample size ratio > 0.10, number of effective samples > 359). For this model, no predictor had a credible effect on LTM accuracy (see Table 2). To investigate between-group memory accuracy independently of attention ratings, we repeated the analysis with only age group as a predictor. This revealed a credible group effect, namely that older adults' memory accuracy was poorer than younger adults' (mean = −0.09; 95% HDI = [−0.14, −0.04]). Effect sizes (standardized coefficients) for these models are reported in Table 2.

As attention ratings were skewed to high responses, we repeated the analyses with binned ratings. The results' pattern did not change either for gibberish detection nor for the LTM analysis (see Appendix F for histograms of response frequencies and supplemental analyses' reports).

3.4. SART

Dependent variables in this task were (1) commission errors and (2) attention ratings. Omissions were not of primary interest but are

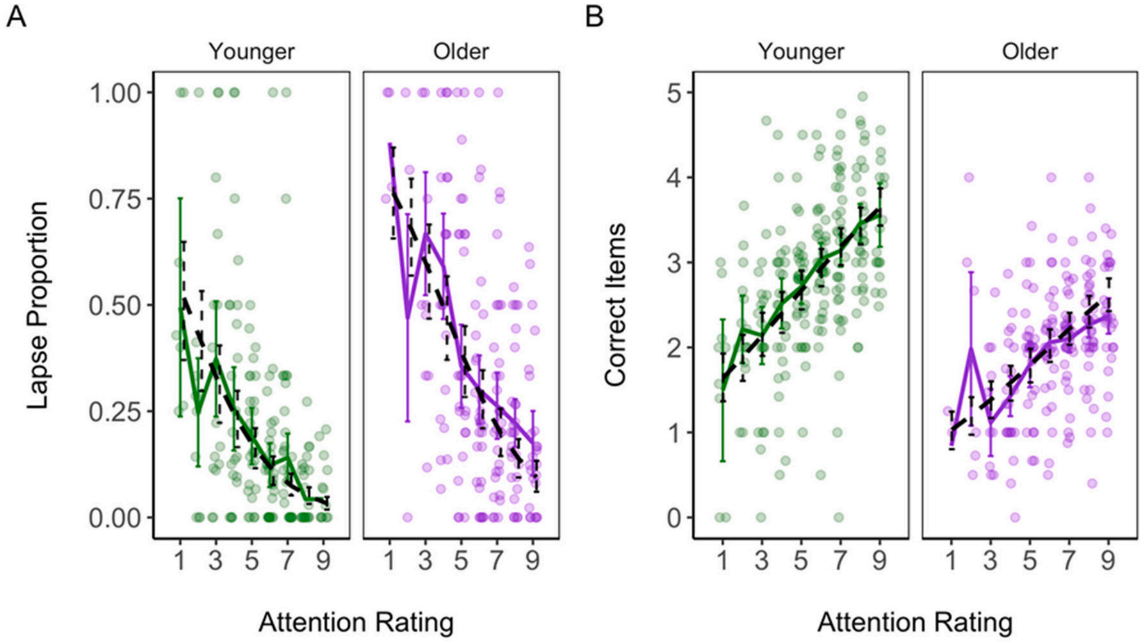

Fig. 3. Proportion of lapse trials (A), and number of correctly reported items (B) in the visual working memory task. Individual points are participants' aggregated data for a given level of the predictor, solid lines are the group mean, and dashed lines are the models' predictions. Error bars represent the 95% within subjects' confidence intervals for the observed data and 95% HDIs for the predicted values.

included in descriptive statistics for completeness (Table 3). Supplementary analyses of reaction times are presented in Appendix B and Fig. D1 in A and Fig. D1 in Appendix D visualizes commission errors and attention ratings as a function of time on task. First, we assessed whether commission errors differed between the attention-probed and the non-probed trials with a 2 (age group) × 2 (probed, non-probed) BANOVA. There was ambiguous evidence against difference between attention-probed and non-probed trials (BF ratio = 0.33), and strong support for an age group difference in error rates (BF ratio = 42), with older adults committing fewer errors. Finally, there was evidence against interaction (BF ratio = 0.20).

It has been suggested that awareness of an error can increase one's attention and hence, an attention probe after an error may not be indicative of one's attention state that led to the commission of the error (Jackson & Balota, 2012; Smallwood, Riby, Heim, & Davies, 2006). Conversely, one could argue that people could infer from having just committed an error that they must have been inattentive, and therefore report less attentiveness when probed right after an error. If the attention reports are influenced by people's errors, rather than the other way around, then we should expect overall lower attentiveness ratings to thought probes after no-go trials than after go- trials, because people commit virtually no errors on the go- trials. There was little evidence for that prediction: Inspection of Table 3 shows that attention ratings were quite similar between no-go and go trials, particularly for older adults. This suggests that errors did not strongly bias attentional reports. Statistically, a 2 (group) × 2 (trial type) BANOVA on attention ratings showed that the best model included the main effect of group and trial type and their interaction (BF_{10} = 3.50). Comparison of the model including trial type to the model without it showed ambiguous evidence for retaining it in the model ($BF_{Group+Type}$ = 0.89 over BF_{Group} = 0.74, BF ratio = 1.20).

Fig. 5 visualizes that commissions decreased as attention ratings increased for both age groups. We assessed this data with a binomial Bayesian hierarchical regression model (adjusted prior scale = [2.84,5.00, 3.57], effective/total sample size ratio > 0.13, number of effective samples > 503). Posterior distributions of the analysis are summarized in

Table 2
Posterior Estimates of the Effect of Predictors of the Visual WM, Gibberish Detection, LTM and SART Analyses.

	Visual working memory task				Reading task				SART	
	Items		Lapses		Gibberish Lapses		LTM Accuracy		Commissions	
Parameters	Mean	95% HDI	Mean	95% HDI	Mean	95% HDI	Mean	95% HDI	Mean	95% HDI
Attention rating	**0.21**	**[0.17, 0.25]**	**−0.43**	**[−0.53, −0.34]**	**−0.36**	**[−0.59, −0.12]**	−0.02	[−0.13, 0.08]	**−0.11**	**[−0.24, −0.005]**
Attention rating†	**0.41**	**[0.33, 0.49]**	**−0.84**	**[−1.03, −0.66]**	**−0.46**	**[−0.76, −0.16]**	−0.02	[−0.17, 0.11]	**−0.20**	**[−0.43, −0.009]**
(2 trials)**	**0.15**	**[0.11, 0.19]**								
(3 trials)***	**0.13**	**[0.10, 0.17]**								
Group	**−0.34**	**[−0.60, −0.06]**	**1.12**	**[0.61, 1.62]**	−0.58	[−2.04, 0.75]	−0.19	[−0.95, 0.52]	−0.48	[−1.13, 0.21]
Group†	**−0.17**	**[−0.30, −0.03]**	**0.56**	**[0.30, 0.81]**	−0.29	[−1.02, 0.37]	−0.09	[−0.54, 0.34]	−0.24	[−0.56, 0.10]
(2 trials)**	**−1.02**	**[−1.16, −0.88]**								
(3 trials)***	**−0.95**	**[−1.09, −0.82]**								
Interaction	−0.06	[−0.14, 0.01]	0.07	[−0.11, 0.27]	0.008	[−0.41, 0.45]	−0.09	[−0.32, 0.13]	−0.07	[−0.30, 0.15]
Interaction †	−0.03	[−0.07, 0.008]	0.03	[−0.05, 0.13]	0.004	[−0.20, 0.22]	−0.04	[−0.18, 0.09]	−0.03	[−0.15, 0.07]
(2 trials)**	−0.02	[−0.08, 0.04]								
(3 trials)***	−0.02	[−0.08, 0.03]								

Note. HDI = Highest Density Interval; Boldface denotes credible effects (parameters' 95% HDI does not include zero). For the visual working memory tasks, supplementary analyses aggregated over two** or three*** trials preceding an attention probe were performed. Parameter values are on the logit scale. † = point summaries of standardized coefficients.

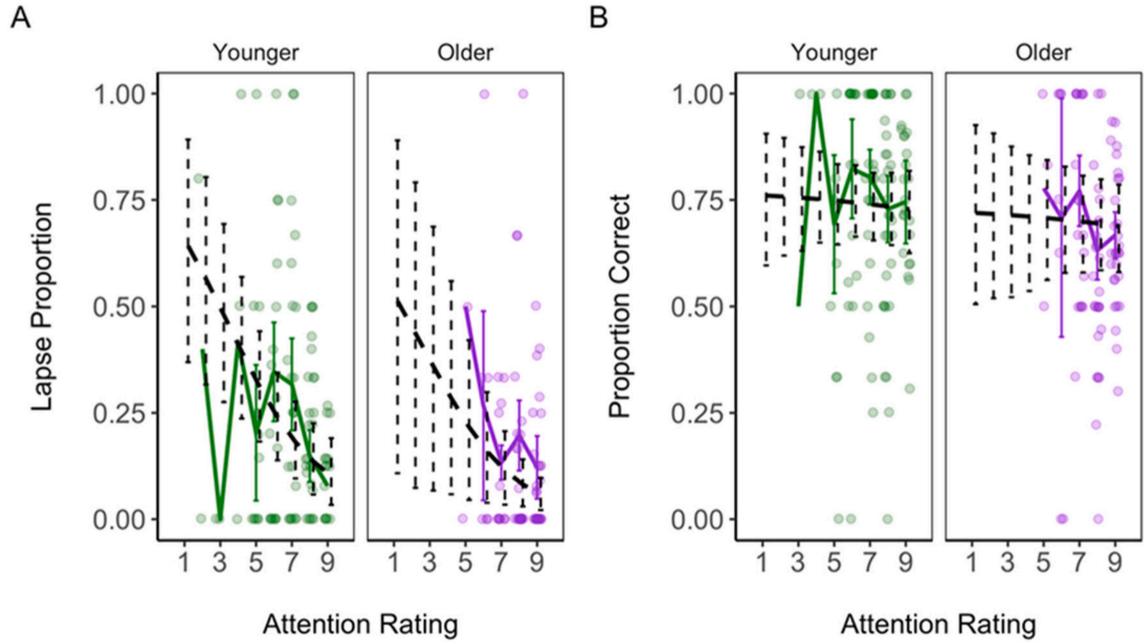

Fig. 4. Proportion of gibberish detection lapses (A), and proportion of correct responses in the long-term memory test (B) in the reading task. Individual points are participants' aggregated data for a given level of the predictor, solid lines are the group mean, and dashed lines are the models' predictions. Error bars represent the 95% within subjects' confidence intervals for the observed data and 95% HDIs for the predicted values.

Table 2. The results showed that higher attention ratings predicted fewer commissions, but age group did not have an additional effect on commission rates. The interaction term was not credible, indicating that attention ratings were similarly predictive of performance in younger and older adults.

As previously, we followed up the lack of age group effect with a repeated analysis, including only age group as a predictor. Here, a credible effect on commission errors was evident (mean = -0.36, 95% HDI = $[-0.59, -0.14]$). Similar to the gibberish analysis, the age-group effect was fully accounted for by the fact that older adults gave higher attention ratings when these ratings were included as predictor in the model. Effect sizes (standardized coefficients) for this model are reported in Table 2.

As attention ratings were skewed to high responses, we repeated the analyses with binned ratings (see Appendix F for histograms of response frequencies and supplemental analyses' reports). This changed the pattern of the results, such that whereas the Group term was still credible, attention ratings were no longer predictive of errors in the SART.

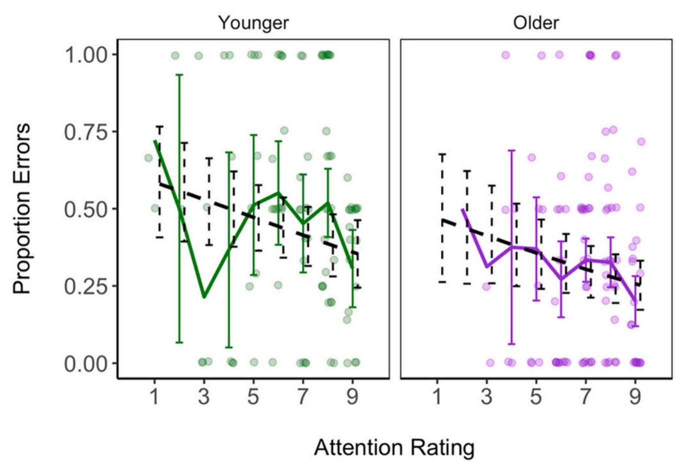

Fig. 5. Proportion of commission errors in the SART task. Individual points are participants' aggregated data for a given level of the predictor, solid lines are the group mean, and dashed lines are the models' predictions. Error bars represent the 95% within subjects' confidence intervals for the observed data and 95% HDIs for the predicted values.

3.5. Perceived task difficulty

One concern for the assessment of self-reported attention across age is that the tasks may present unequal difficulty for older compared to younger adults, which may lead to a bias in self-rated attention. To monitor perceived difficulty across tasks and age, at the end of each task, we solicited ratings on a 1–9 scale (1 = "*Not at All Difficult*", 9 = "*Very Difficult*"), visualized in Fig. 6A. Submitting these ratings to a 2 (age group) × 3 (task) BANOVA showed that best model included the main

Table 3
Means and Standard Deviations (in Parentheses) of the Dependent Variables in the SART task.

Variable	Probed trials		Non-probed trials	
	Younger	Older	Younger	Older
Commissions	0.43 (0.18)	0.29 (0.22)	0.45 (0.16)	0.31 (0.18)
Omissions	0.01 (0.03)	0.004 (0.02)	0.009 (0.01)	0.01 (0.01)
Attention ratings (no-go)	7.05 (1.58)	7.32 (1.21)	–	–
Attention ratings (go)	6.73 (1.73)	7.34 (1.27)	–	–
Attention rating reliability	0.85	0.82	–	–

Note. Commissions (no-go errors) and omissions (go-errors) are reported as a proportion. Reliability estimates were calculated with the Spearman-Brown's correction.

effects of group and task ($BF_{10} = 1.56 \times 10^{11}$). Models comparisons yielded strong evidence for the effect of task ($BF_{Group+Task} = 1.56 \times 10^{11}$ over $BF_{Group} = 0.70$, BF ratio = 2.13×10^{11}). Older adults reported somewhat higher difficulty, but there was no credible evidence for (or against) the effect of age, as the evidence was ambiguous ($BF_{Group+Task} = 1.56 \times 10^{11}$ over $BF_{Task} = 1.43 \times 10^{11}$, BF ratio = 1.04). The interaction of these factors was rejected (BF ratio = 0.10).

3.6. Attention and current concerns

Descriptive statistics as well as split-half reliabilities of attention ratings in the visual working memory and reading tasks are presented in Table 1, and these measures for the SART appear in Table 3. Fig. 6B presents the average ratings of attention per age group and task.

First, we checked whether attention ratings differed as a function of age group and task with a 2 (age group) × 3 (task) BANOVA. Overall, attention reports were higher for older compared to younger adults, and they were higher in the reading task than the SART task, with the lowest values being observed in the visual working memory task. Accordingly, the best model in the BANOVA included only the main effects of age group and task ($BF_{10} = 2.65 \times 10^9$), which was preferred by a factor of 8 to the second-best model, which includes the interaction.

The second goal of this study was to investigate whether (a) younger and older adults differ in the number of current concerns or concern importance ratings, as suggested by the *Control Failure x Concerns* view of MW (McVay et al., 2013), and (b) whether these variables predict attention levels during the tasks. Dependent variables were (1) the number of concerns and (2) mean concern importance rating per participant. Two older adults' questionnaires were not completed according to the instructions and not analyzable. A concern was coded whenever a statement was formulated in reference to the participant (e.g., "I want to do more sports"). No concern was coded if the statement was not self-referent (e.g., "Sports are a good thing"). Importance ratings were only included in the analysis if a concern was coded. Concerns listed twice within the questionnaire (e.g., "I want to do more sports" both in the categories "health" and "fitness") were coded only once.

Descriptive statistics are summarized in Table 4. Bayesian *t*-tests showed ambiguous evidence against an age group difference in the sum of concerns ($BF_{10} = 0.35$), and also in mean concern importance ($BF_{10} = 0.44$). Next, we assessed whether age group and sum of participants' current concerns predicted mean attention ratings in Bayesian linear regressions implemented in rstanarm. We used the default priors provided for the Gaussian family (adjusted scale = 8.58 for the intercept and = [2.14,2.38,2.61] for the coefficients) with the identity link and report the full model as the most informative. The number of effective samples was > 1549 and the effective/total sample size ratio > 0.58. The model's posterior parameters are summarized in Table 5 and visualized along the data in Fig. 7. The main effect of age group indicated that older adults' overall attention ratings were higher, but there was no credible effect of sum of current concerns on attention ratings, nor an interaction. The results did not change with the mean of concerns as a predictor.

Next, we repeated the analysis with mean rated concern importance (adjusted prior scale = 8.58 for the intercept, and [2.14,1.08,0.72] for the coefficients, effective samples > 1617, effective/total sample size ratio > 0.43). The results followed the same pattern: older adults' mean rated attention was higher than younger adults', but mean concern importance did not predict attention ratings.

4. Discussion

The current study assessed the validity of older adults' attention reports in comparison to younger adults. We measured attention state on a continuous scale to circumvent problems of interpretation and criterion changes in classifying thoughts in dichotomous categories across age. Thought probes were presented in two tasks that were previously employed in MW research, namely a reading task and SART, as well as in a visual working memory task.

4.1. Are the on-task reports of older adults valid?

Our main concern was to understand whether older adults' attention reports covary with fluctuations in task performance, as has been previously observed for younger adults in working memory tasks (Adam & Vogel, 2017; Krimsky et al., 2017; Unsworth & Robison, 2016) as well as in MW studies using a continuous scale to measure on-task attention (Allen et al., 2013; Levinson et al., 2012; Mittner et al., 2014; Morrison, Goolsarran, Rogers, & Jha, 2014; Qin, Perdoni, & He, 2011; Ruby, Smallwood, Engen, & Singer, 2013; Smallwood, Ruby, & Singer, 2013). The covariation between self-reported fluctuations in attention and fluctuations in performance has been taken as evidence of good metaknowledge about one's attentional state (Adam & Vogel, 2017). Our rationale was that observing these covariations would allow us to assess how well older adults can monitor the degree to which their attention is currently on task.

We obtained largely consistent results across all three tasks: There was a credible relation between attention ratings and task performance

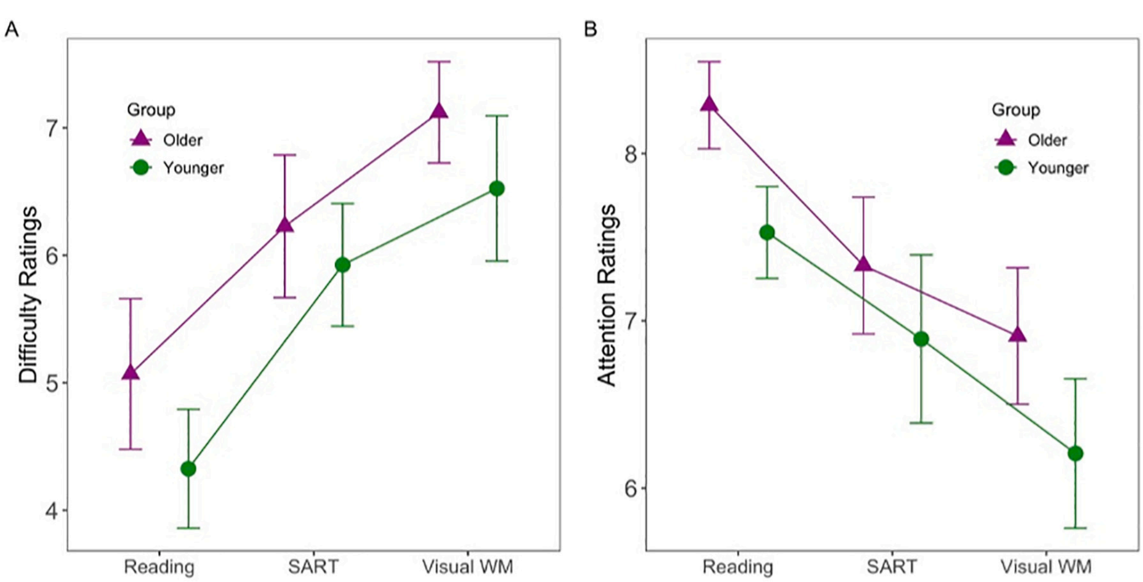

Fig. 6. Difficulty and Attention ratings per task. Error bars represent the 95% within-subjects' confidence intervals. Note that difficulty ratings were solicited once per task, whereas attention was surveyed multiple times in each task.

Table 4
Descriptive Statistics for Current Concerns and Concern Importance Ratings per Life Area and Age Group.

	Sum of concerns (range)		Mean of concerns (*SD*)		Mean importance ratings (*SD*)	
Area	Younger	Older	Younger	Older	Younger	Older
Budget	50 (0–4)	38 (0–3)	1.25 (0.89)	0.90 (0.93)	6.01 (3.51)	4.50 (4.17)
Fitness	56 (0–5)	49 (0–5)	1.40 (1.03)	1.16 (1.34)	5.62 (2.91)	4.68 (3.97)
Health	56 (0–4)	57 (0–6)	1.40 (0.90)	1.35 (1.24)	6.68 (3.24)	6.32 (3.84)
Home	73 (0–4)	62 (0–6)	1.82 (1.10)	1.47 (1.58)	6.45 (2.65)	5.14 (3.82)
Other	11 (0–5)	35 (0–5)	0.27 (1.01)	0.83 (1.24)	0.83 (2.56)	3.50 (4.32)
Recreation	74 (0–4)	70 (0–5)	1.85 (1.09)	1.66 (1.37)	6.69 (3.02)	6.88 (3.32)
Relationships	67 (0–5)	60 (0–7)	1.67 (1.20)	1.42 (1.62)	6.82 (3.33)	5.96 (4.06)
Work	67 (0–5)	39 (0–4)	1.67 (1.16)	0.92 (1.23)	6.82 (3.09)	3.92 (4.26)
Self-Development	74 (0–4)	70 (0–5)	1.85 (1.09)	1.66 (1.37)	6.69 (3.02)	6.88 (3.32)
Overall	528	480	1.41 (0.74)	1.21 (1.02)	5.74 (1.48)	5.09 (2.44)

Note. SD = Standard Deviation.

Table 5
Posterior Estimates of the Effect of Number (Sum) of Current Concerns and Mean Concern Importance on Attention Ratings.

Parameters	Sum of Concerns		Concern Importance	
	Mean	95% HDI	Mean	95% HDI
Intercept	7.03	[6.70, 7.35]	6.75	[6.15, 7.35]
Age Group	**0.72**	**[0.05, 1.38]**	**1.16**	**[0.05, 2.29]**
Concerns	0.01	[−0.009, 0.03]	0.07	[−0.02, 0.17]
Age Group x Concerns	−0.004	[−0.04, 0.04]	−0.08	[−0.27, 0.10]

Note. HDI = Highest Density Interval; Boldface denotes credible effects (parameters' 95% HDI does not include zero).

in both age groups. For all tasks, there was no evidence for an interaction between attention rating and age group, indicating that attention reports were equally predictive of performance for younger and older adults. This finding supports the validity of attention self-reports across age. The only exception to this general pattern was performance in the LTM test for the reading passages, which did not yield credible evidence for any relation between attention ratings and performance in either age group. We surmise that this null result occurred because we tested memory for incidentally encoded details that were unimportant for the text comprehension or gibberish detection. Hence, it is likely that our attention assessment, targeting attention to the task, did not capture attention to these relatively task-irrelevant details, and was therefore unrelated to how well these details were encoded into episodic memory.

Overall, our results suggest that healthy older adults' ability to monitor and report their current degree of attention is comparable to that of younger adults'. These results resonate with studies showing preserved meta-cognitive monitoring in aging (Castel, Middlebrooks, & McGillivray, 2015; Hertzog & Dunlosky, 2011). These results are also consistent with a recent study investigating variations in visual working memory performance and meta-cognitive monitoring across age in a big sample of over 600 individuals (Mitchell, Cusack, & Cam-CAN, 2018). Mitchell and Cusack showed that older adults tended to be slightly overconfident in low memory load trials, but reports tracked trial-by-trial performance fluctuations similarly across age. Furthermore, better cognitive monitoring was associated with better performance irrespective of age. Altogether, these findings and ours point to the possibility that the intercept of older adults' attention ratings may be shifted upwards, but the attention-performance relationship can be described with the same slope in both age groups.

4.2. Why do older adults report to be on task more than younger adults?

We addressed a concern raised by Frank et al. (2015) that younger and older adults might differ in their criterion for classifying attention focus using dichotomous categories such as on-task vs. MW. Previous research has shown that older adults report fewer MW than younger adults and higher proportion of task-related interference thoughts (Frank et al., 2015; McVay et al., 2013; Zavagnin et al., 2014), but the reasons for these differences are yet under debate. Here, we used a continuous rating scale to minimize the possibility that criterion

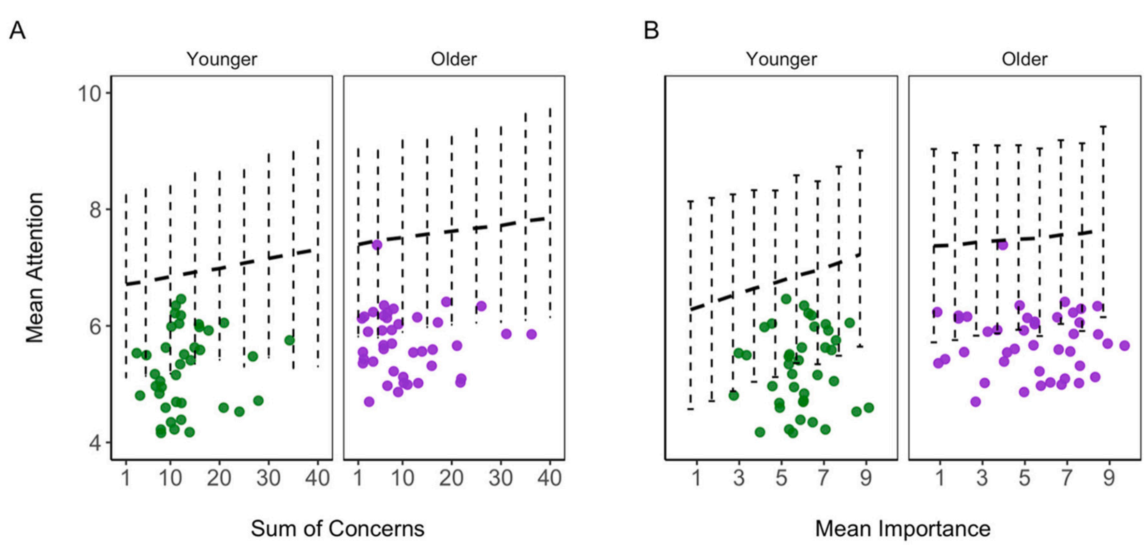

Fig. 7. The scatterplots display individual participants' responses in the concerns questionnaire (x-axis) and their mean attention rating across the tasks (y-axis) for number of concerns (left panel) and rated concern importance (right panel). Dashed lines are the regression models' predictions. Error bars are the 95% HDIs for the predicted values.

changes between groups in classifying thoughts in one category vs. the other would obscure the assessment of the relationship between self-reports and performance. The use of a continuous scale did not change the main pattern of findings in the literature: Older adults' overall reported attention was higher than younger adults'.

Our results are consistent with two possibilities. One is that older adults have a bias to report higher attention; the other is that they actually sustain their attention focus more often or more continuously on the current task than younger adults do. For the gibberish detection and SART tasks we have no reason to believe that older adults' attention ratings were biased upward: Attention ratings predicted performance in both age groups with the same regression line. In the visual working memory task, older adults' higher attention ratings were accompanied by poorer performance, so that using the same regression line for both age groups would lead to an over-prediction of older adults' performance. This could mean that the attention ratings of older adults were biased upward. Alternatively, it could mean that their performance is impaired relative to that of younger adults' by a factor independent of their attentional state, such that, with an equal degree of attentional focus on the task, older adults perform more poorly than younger adults.

One possibility that could lead to an upward attention-rating bias is that the tasks may be more difficult for the older individuals. Given that VWM capacity is reduced in older age (Brockmole & Logie, 2013), this concern is especially plausible for the VWM task. Although participants' ratings of perceived task difficulty did not differ across age in the present study, it is of theoretical importance to consider the implications of task demand on the experience and self-appraisal of attention. VWM maintenance is assumed to demand attention (C. C. Morey & Bieler, 2013), and some theoretical accounts view VWM as sustained attention to the memory representations (Chun, 2011). Accordingly, the difficulty to maintain memory representations accessible may lead to higher attention investment and hence, to valid ratings of higher attention.

An alternative possibility is that perceived task difficulty could serve as an internal cue for attention appraisals, without leading to increased attentional effort. According to the cue-utilization approach to meta-memory (Koriat, 1997), participants use a variety of cues, − in addition to evaluating the strength of a memory trace – when forming predictions of their own memory performance on a delayed test. On a more difficult task, the effort of memory maintenance could thus serve as a heuristic cue for attention appraisal, thereby resulting in upward-biased ratings.

However, previous research showed that higher attentional demand tends to decrease self-reported attentiveness. Higher task difficulty is assumed to demand more attention control, and this construct has been linked to individual differences in MW propensity in younger adults (Kane et al., 2007; McVay & Kane, 2012b). According to these studies, younger individuals with lower WM capacity self-report lower attention during more demanding tasks. This resonates with our findings that self-rated attention was higher in tasks rated as less difficult for both age groups. Therefore, we find it implausible that older adults' attention ratings were biased upward by experiencing the tasks as more difficult than young adults, and we find it even more implausible that the attention ratings are biased upward in one task (visual WM) but not two others.

Conversely, on the assumption that older adults' working memory capacity is much reduced compared to that of young adults', it is plausible that their performance on a test of working memory is impaired, whereas their performance on two other tasks that rely much less on working memory is not impaired. This interpretation implies that older adults' reduced working memory capacity does not arise from a higher prevalence of attentional lapses in that age group. Rather, it arises from a genuine reduction in their ability to hold information in working memory, for instance due to a reduction in memory resources (Ma, Husain, & Bays, 2014) or due to an increased vulnerability to interference (Oberauer & Lin, 2017).

Assuming that the attention ratings of older adults are valid, we must conclude that they do invest more sustained attention than younger adults, either due to higher motivation (Staub, Doignon-Camus, Bacon, & Bonnefond, 2014; Tomporowski & Tinsley, 1996) or as a compensation for age-related decline in ability (Staub et al., 2014). The results regarding age differences in performance-based measures of sustained attention are mixed: Some investigations report that older adults' sustained attention performance is on par or even superior to that of younger adults' (Carriere, Cheyne, Solman, & Smilek, 2010; Jackson & Balota, 2012; Jackson, Weinstein, & Balota, 2013; Parasuraman, Nestor, & Greenwood, 1989; Staub et al., 2014; Tomporowski & Tinsley, 1996), whereas other studies found a decline in older adults' sustained attention compared to younger participants (Fortenbaugh et al., 2015; Giambra, 1997; McAvinue et al., 2012). For example, assessing > 600 individuals aged 14 to 77 on the SART, Carriere et al. (2010) found linear decreases in commission errors with age, implying that sustained attention got better with advancing age. By contrast, another large online study observed an age-related decline in sustained attention with $N > 10.000$ (Fortenbaugh et al., 2015). However, the sample size in this study was heavily skewed towards younger age, such that the number of 60+ participants was disproportionately small, and the task duration was rather short (four minutes). Therefore, it is unclear whether older adults may have been somewhat disadvantaged by the short time to adjust to the task. Recent work suggested that the discrepant results in the literature may be due to whether the task activates automatic processes (as the regular SART, requiring frequent responses) or controlled processes, as in a vigilance task with responses to rare targets (Staub, Doignon-Camus, Marques-Carneiro, Bacon, & Bonnefond, 2015). In a lab-based study, these authors found that age effects in the tasks were reversed: Older adults showed decrements in sustained attention when rare, but not when frequent responding was required, whereas the pattern for younger adults was the opposite. They suggested that older adults are able to sustain their attention even better than younger adults because they adopt a more controlled strategy, but this approach is beneficial only for tasks with a high level of automaticity (Staub et al., 2015), such as the SART that is widely used in MW research. We surmise that older adults may deploy sustained attention more consistently than younger adults as a strategy to compensate for decline in working memory capacity, which is greatly reduced in older adults (Brockmole & Logie, 2013).

In contrast to the *Control* Control*Failure x Concerns* theory (McVay & Kane, 2010), Smallwood and Schooler (2006) proposed that MW itself requires executive resources, supporting their argument with the evidence that off-task thought is reduced when task demands increase. These authors suggested that, as MW episodes reflect a coordinated, coherent train of thought, executive control is required to coordinate information to enable this mentation. The prediction that may be derived from this view is that older adults mind-wander less than young adults, as their reduced working memory capacity does not allow to divide attention between on- and off-task thoughts as well as it does for younger adults. In a recent theoretical development, Smallwood (2013) further asserted that executive attention is specifically needed for the continuation of engagement in MW to ensure the continuity of internal thought. If this is the case, it is possible that, due to age-related decline in executive resources, older adults both lapse more frequently and are also less able to sustain the off-task train of thought, their attention being rapidly returned to the task due to its external salience, as older adults rely more on external information (Lindenberger & Mayr, 2014). As a consequence, their mind would stray off task more often, but only very briefly, so that a larger proportion of attention probes catches them during on-time episodes. In conclusion, the hypothesis of Smallwood and Schooler is better compatible with our findings than the *Control Failure x Concerns* theory.

4.3. MW and current concerns

Contrary to Parks et al. (1989), we did not observe an age difference in personal concerns: the comparisons yielded ambiguous evidence. Furthermore, concerns did not predict mean reported attention. Older adults may nurture fewer long-term goals, as their future entails fewer normative life events – biological, psychological and social life transitions that are highly correlated with chronological age (Cavanaugh & Blanchard-Fields, 2018) – than younger adults'. However, healthy and active older adults probably have a busy agenda that is comparable to that of younger adults', and thus a commensurate amount of current concerns.

The lack of correlation between concerns and mean attention does not disprove the contention that MW emerges from the interaction of current concerns, environmental cues, and the person's attention control ability (McVay et al., 2013). Rather, it suggests a methodological challenge: Possibly, the relationship between MW and concerns becomes evident across a longer time span than that of a lab-based study. For example, Song and Wang (2012) investigated MW with daily-life experience sampling, collecting detailed descriptions of MW episodes. When participants reported that an occurrence of MW was triggered by an internal cue (a thought or a memory), the content of these MW episodes was significantly related to these participants' personal plans – in other words, their current concerns. Thus, assessing the relationship between concerns, MW occurrences and executive control as proposed by the *Control Failure x Concerns* view may require a combination of quantitative and qualitative data collected across longer time periods.

Another possibility is that the relationship between current concerns and MW manifests on a different time frame than is assessed by the questionnaire we used, which inquiries about long-term, high-importance life goals. In contrast, recent research revealed that the contents of future-oriented MW often revolve around short-term goals such as errands, which may be described as a prospective memory function (for a review, see Kvavilashvili & Rummel, 2020). Indeed, a large body of studies addressing qualitative aspects of the MW reports showed that MW thoughts are often future-oriented (for a review, see Stawarczyk, 2018). This finding has been also reported in a sample of older adults (Jackson et al., 2013). Moreover, Warden, Plimpton, and Kvavilashvili (2019) found that younger and older adults reported thinking about the future equally frequently. Future-thought contents often concern immediately pending activities (Baird, Smallwood, & Schooler, 2011) and tasks upcoming within the next month (Plimpton, Patel, & Kvavilashvili, 2015), week, or even on the same day (Baumeister, Hofmann, Summerville, Reiss, & Vohs, 2020; Stawarczyk, Cassol, & D'Argembeau, 2013; Stawarczyk, Majerus, Maj, Van der Linden, & D'Argembeau, 2011). Future-oriented spontaneous thoughts may serve the function of prospective memory rehearsal and thus help achieve peoples' daily goals (Klinger, 2009; Kvavilashvili & Rummel, 2020; Mooneyham & Schooler, 2013), as the frequency of spontaneous future-oriented thoughts positively correlates with successful prospective memory-task completion (Seli, Smilek, Ralph, & Schacter, 2018) and memory recall (Steindorf & Rummel, 2017).

Together, this evidence indicates that current concerns – particularly, imminent action goals – play a role in MW. Therefore, it may be important to survey participants' near-future goals, and collect this data on each experimental session. For example, Parks et al. (1989) administered all questionnaire measures in a single session, whereas we did not control for questionnaire filling time, encouraging participants to take the questionnaires home if they wished. In the autobiographical memory literature, it has been observed that older adults tend to generate less detailed episodic events in a given time interval compared to younger adults (Addis, Musicaro, Pan, & Schacter, 2010; Addis, Wong, & Schacter, 2008). Although also some younger adults took questionnaires home, we surmise that especially for older adults, the extra time may have enabled a more complete listing of their relevant concerns.

In any case, we found no evidence for the contention that the reason for older adults' reduced report of MW is their decreased number or importance of current concerns.

5. Conclusion

Older adults report less MW episodes than younger adults. This challenges the view of MW as arising from failures of executive control, an ability assumed to decline with aging. Here we assessed and found no evidence for the hypothesis that older adults are less able to appraise fluctuations in their attentional state leading to low validity of their self-reports of MW. Our results point to a preserved ability to monitor one's own state of attentional focus in older age.

This means that reduced reports of MW in older age are probably valid. We also found little evidence for a systematic, task-general bias to underreport MW in older compared to younger adults. Our results underscore the importance of empirically addressing both components of the *Control Failure x Concerns* view (McVay & Kane, 2010), to promote the development of theoretical explanation of the mind-wandering experience.

Appendix A. Modified personal aspirations and concerns inventory

The questionnaire was administered in German. The instruction read:

“Thank you in advance for completing this questionnaire. Your information will be treated confidentially and analyzed in anonymized form.

Instruction. A current concern is a topic that presently occupies you. Undoubtedly, you have concerns in different areas of your life. Probably you think about how to solve these concerns. A concern is not only understood as a problem. You could have concerns about unpleasant things which you would like to get rid of, prevent or avoid. Or you could have things that you'd like to obtain, preserve, or complete. Concerns are also thoughts that often come to your mind because you find them interesting or important.

There are 9 life areas: (1) Home / Household, (2) Job / Occupation, (3) Education / Personal Development, (4) Leisure / Recreation, (5) Relationships, (6) Health, (7) Fitness / Sports, (8) Budget, and (9) Other. Please name concerns you feel you currently have in these areas and rate the importance of each concern to you. You can list as few or as many concerns per area as you feel you have. Please do not hesitate to contact the research assistant if you have any questions or comments.” Table A1 depicts an example question as presented in the questionnaire.

Table A1
Example Question from the Concerns Questionnaire Used in the Study. Each Question Provided Ten Response Rows.

1) Do you have current concerns in the area Home / Household?	If 0 = barely important, 5 = moderately important and 10 = highly important, how important is each concern to you?

(*continued on next page*)

Table A1 (*continued*)

1) Do you have current concerns in the area Home / Household?	If 0 = barely important, 5 = moderately important and 10 = highly important, how important is each concern to you?

Appendix B. SART reaction time analyses

Here, we report descriptive statistics and BANOVAS for $N = 56$ (22 younger, 34 older adults). Due to a programming error, reaction times (RTs) were not recorded for 32 participants. The summarized RTs in milliseconds were $M_{GoTrials} = 296$ ($SD_{GoTrials} = 15$) and $M_{NoGo} = 273$ ($SD_{NoGo} = 15$) for younger adults, and $M_{GoTrials} = 332$ ($SD_{GoTrials} = 25$) and $M_{NoGo} = 331$ ($SD_{NoGo} = 25$) for older adults. The data are visualized in Panel A of Fig. B1.

B.1. Analysis A

We conducted a 2 (Age Group: Younger, Older) × 2 (Trial Type: Go, No-Go) BANOVA on mean RTs. Results are presented in Table B1, column A. There was largest support for the model with main effects and their interaction. However, the contribution of the interaction was not credible (BF in favor of keeping the interaction in the model = 1.46). Follow-up *t*-tests indicated that the interaction term was ambiguous due to the fact that older adults' RTs were similar across trial types (BF = 0.25), but younger adults' No-Go RTs were faster than their RTs on Go-trials (BF = 3.34). This indicates that younger adults' RTs were faster when they made a commission error (i.e., responded in no-go trials).

Table B1
Bayes Factors (BFs) for the 2 × 2 Bayesian ANOVAS on Reaction Time.

	Predictor (2)			
	A	B	C	D
Predictors included in the model	Trial Type (go, no-go)	Outcome (Commission, Correct Rejection)	Outcome (Commission, Correct Rejection)	Outcome (Commission, Correct Rejection)
(1) Age Group	5×10^5	574	**1.20**	1.48
(2) See column	0.88	148	0.24	3.13
(1) + (2)	4.86×10^5	$\mathbf{8.37 \times 10^4}$	0.31	**4.79**
(1) + (2) + (1) × (2)	$\mathbf{7.12 \times 10^5}$	6.66×10^4	0.75	2.25

Note. Predictor (2) is indicated above each column. (1) + (2) = model with both main effects, (1) + (2) + (1) × (2) = model with main effects and their interaction. Dependent variables in the analyses were: A = mean RT across conditions, B = mean RT 3 trials prior to a no-go target, C = RTCV on 3 trials prior to no-go target, D = RT one trial after a no-go target. The best model for each column is printed in bold. The BFs represent evidence for each model against the model with no fixed effects, and only a random effect of subjects.

B.2. Analysis B

Next, we analyzed RTs on the three trials preceding a no-go target as a function of the no-go outcome (correct rejection or commission error). The data were trimmed to include only no-go trials preceded by three consecutive go-trials, and 69% of the data survived the trimming procedure. The trimmed data was balanced between subjects, as all participants experienced the identical sequence of SART stimuli resulting in the exclusion of the same trials for everyone. Small variations in the number of observations resulted from the exclusion of observations with any omissions in the preceding three trials.

The data are visualized in Panel B of Fig. B1 and summarized in Table B2. we conducted a 2 (Age Group: Younger, Older) × 2 (Outcome: Commission, Correct Rejection) BANOVA (see results in Table B1, column B). This revealed strongest support for a model with only the main effects of age group and outcome, indicating that (1) older adults' responses were slower overall, and (2) commission errors were preceded by faster mean RTs than correctly withheld responses in both age groups.

Table B2
Means and Standard Deviations (in Parentheses) for RTs (in Milliseconds) in the SART task.

Variable	Commission		Correct	
	Younger	Older	Younger	Older
Mean 3 trials before target	286 (15)	328 (20)	313 (15)	340 (20)
RTCV 3 trials before target	0.28 (0.03)	0.30 (0.08)	0.24 (03)	0.34 (0.08)
RT 1 trial post-target	291 (17)	302 (31)	298 (17)	323 (31)

B.3. Analysis C

The following analysis relates RT variability to no-go trial outcome. Variability was computed as coefficient of variation (RTCV), namely

$SD_{GoTrials}$ / $Mean_{GoTrials}$ on the three trials preceding a no-go target, using the trimmed data described in the previous analysis. Descriptive statistics are summarized in Table B2, and data are visualized in Panel C of Fig. B1. A 2 (Age Group: Younger, Older) × 2 (Outcome: Commission, Correct Rejection) BANOVA (see Table B1, column C) yielded low to no evidence for all models.

B.4. Analysis D

Lastly, we addressed post-error slowing. The data are summarized in Table B2 and visualized in Panel D of Fig. B1. A 2 (Age Group: Younger, Older) × 2 (Outcome: Commission, Correct Rejection) BANOVA on go-RTs one trial after a no-go target showed that the best model included the main effects of age group and outcome (see Table B1, column D). Parsing of the evidence in BF ratios revealed support for the factor Outcome (BF ratio = 3.22) but not Age Group (BF ratio = 1.52), indicating that post-error slowing was present for both younger and older adults.

In summary, (1) older adults' RTs were slower than younger adults', and (2) commission responses were preceded by faster responses on go-trials. These findings replicate previous reports. In contrast, (3) RT variability did not predict no-go trial outcome, and (4) there was some evidence that RTs on one trial after a correct response were slower than ones after a commission error.

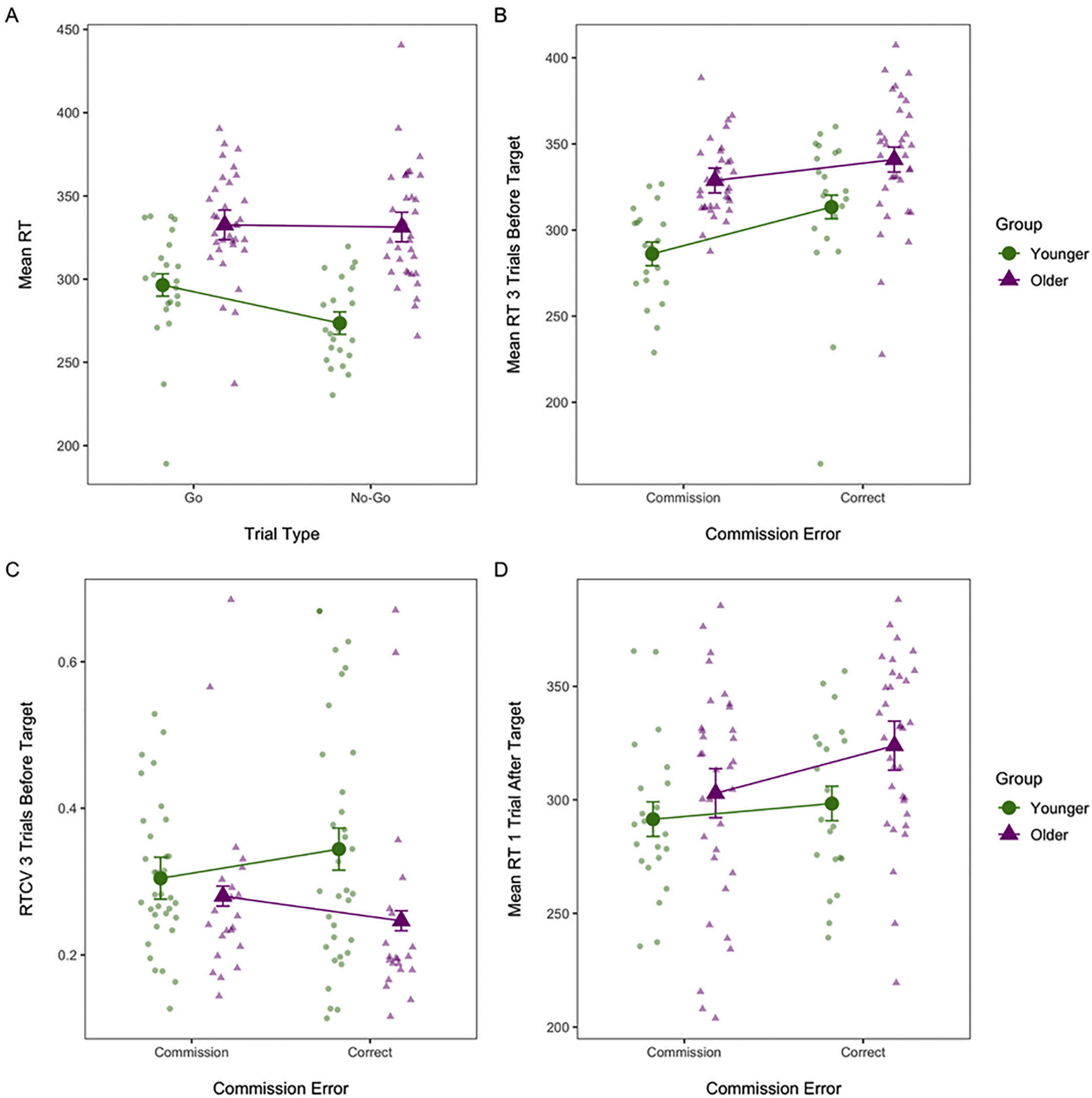

Fig. B1. Visualization of the SART RT data. Error bars are the 95% within-subjects' confidence intervals.

Appendix C. Evaluation of differing task lengths on attention across age

A possible challenge to our study is that each task used different numbers of trials and led to different overall task lengths. Furthermore, as older adults show slower reaction times in self-paced tasks such as visual WM (Duarte et al., 2013; Strunk, Morgan, Reaves, Verhaeghen, & Duarte, 2018) and reading (Jackson & Balota, 2012; Krawietz et al., 2012), their on-task time is bound to be longer. In contrast, the automatized SART pace used here ensured invariant task duration across groups. Critically, using 300 trials in the visual WM task added 30 min to the session length for older adults. One may wonder whether on-task duration impacted self-reported attention.

C.1. Visual WM task

To examine the effect of task length on attention, we divided the visual WM task in 6 subblocks and compared attention ratings as a function of block. The data are visualized in Fig. C1, Panel A. A 2 (Age Group: Younger, Older) × 6 (Block) BANOVA showed largest support for the model with main effects and their interaction (see Table C1). The evidence for including the interaction in the best model was BF = 15 (BF_{Full} divided by $BF_{MainEffects}$). However, it was not older adults' attention that dwindled: Follow-up *t*-tests of mean attention between the first and sixth blocks showed credible declines in younger adults' rated attention ($BF_{10} = 1.68 \times 10^{12}$), but not in older adults' ($BF_{10} = 0.48$). For younger adults, mean attention declined by one rating point between the first and the last block ($M_{Block1} = 6.78$, $SD = 1.42$, $M_{Block6} = 5.62$, $SD = 1.90$). For older adults, this decrease was of only half a rating point ($M_{Block1} = 7.21$, $SD = 1.54$, $M_{Block6} = 6.74$, $SD = 1.75$).

Table C1
Bayes Factors (BFs) for the 2 × 2 Bayesian ANOVAS on Attention Ratings and Performance Lapses.

	Task			
	Visual WM		Reading	
Predictors included in the model	Attention ratings	Performance lapses	Attention ratings	Detection lapses
(1) Age Group	2.46	49	116	6.49
(2) Block	1.29×10^{11}	4.76×10^{11}	9.30×10^{4}	3.21×10^{20}
(1) + (2)	3.18×10^{11}	2.10×10^{13}	1.11×10^{7}	2.40×10^{21}
(1) + (2) + (1) × (2)	$\mathbf{4.97 \times 10^{12}}$	$\mathbf{4.55 \times 10^{14}}$	$\mathbf{5.15 \times 10^{8}}$	$\mathbf{1.03 \times 10^{23}}$

Note. Dependent variable is indicated above each column. (1) + (2) = model with both main effects, (1) + (2) + (1) × (2) = model with main effects and their interaction. The best model for each column is printed in bold. The BFs represent evidence for each model against the model with no fixed effects, and only a random effect of subjects.

Next, we analyzed the relationship of time on-task and proportion of performance lapses (Fig. C1, Panel B). As in the main analysis, a lapse was defined as a trial with less than 2 correctly reported items. A 2 (Age Group: Younger, Older) × 6 (Block) BANOVA on proportion of lapses showed largest support for the model with the main effects and their interaction (see Table C1). Follow-up *t*-tests revealed that the interaction was due to older adults' performance improving towards the end of the task ($BF_{10} = 2526$) with less lapses in Block 6 ($M = 0.16$, $SD = 0.17$) than in Block 1 ($M = 0.39$, $SD = 0.24$). For younger adults, there was ambiguous evidence for a difference in performance between Block 1 ($M = 0.2$, $SD = 0.18$) and Block 6 ($M = 0.11$, $SD = 0.12$), $BF_{10} = 2.29$.

Next, we analyzed attention ratings in the reading task, similarly dividing the data into six blocks, and observed the same pattern. The data are visualized in Fig. C1, Panel C. A 2 (Age Group: Younger, Older) × 6 (Block) BANOVA showed that the full model was the best model, ($BF_{10} = 5.15 \times 10^{8}$), and there was strong evidence for including the interaction term in this model (BF ratio = 46). WW WWithin-age group *t*-tests showed that whereas evidence for a change between the first and last block in the younger group was ambiguous (BF = 1.15), the evidence was against such a change for older adults (BF = 0.28), indicating stability of rated attention across the task's length. Thus, despite the longer durations, older adults actually upheld their attention for the whole session, indicating that greater task lengths were not a confound for the validity of the attentional reports between groups.

Lastly, to test the relationship of performance and time on task, we analyzed the proportion of gibberish detection lapses with a 2 (Age Group: Younger, Older) × 6 (Block) BANOVA. The data are visualized in Fig. C1, Panel D. The results again supported the model with the main effects and their interaction (see Table C1), and BF in favor of the interaction was 42. Within-group *t*-tests showed mild evidence for a change from block one to six for younger adults in detection of gibberish ($BF_{10} = 3.65$) and evidence against a change for older adults ($BF_{10} = 0.27$).

In summary, analyses of attention ratings as well as performance accuracy did not reveal disadvantages to older adults as a consequence of longer time on task.

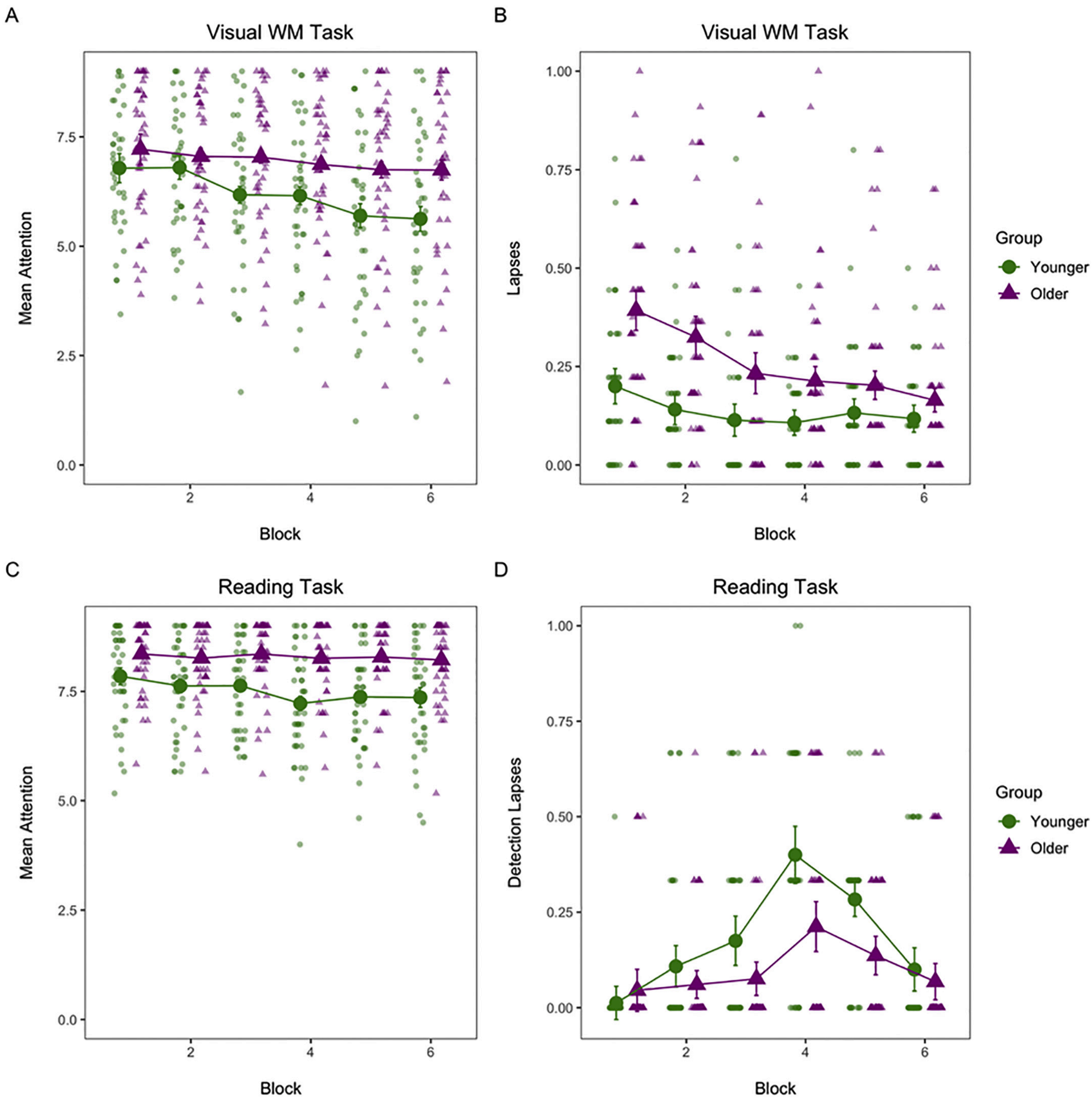

Fig. C1. Visualization of mean attention ratings and performance lapses (y-axes) as a function of time on task (x-axis) in the visual WM (Panels A and B) and reading (Panels C and D) tasks. Error bars are the 95% within-subjects' confidence intervals.

Appendix D. Visualization of commission errors and mean attention Ratings throughout the SART

To complement the visualizations of task accuracy and mean attention ratings across time on task in Appendix C, Fig. D1 presents the same plots for the SART.

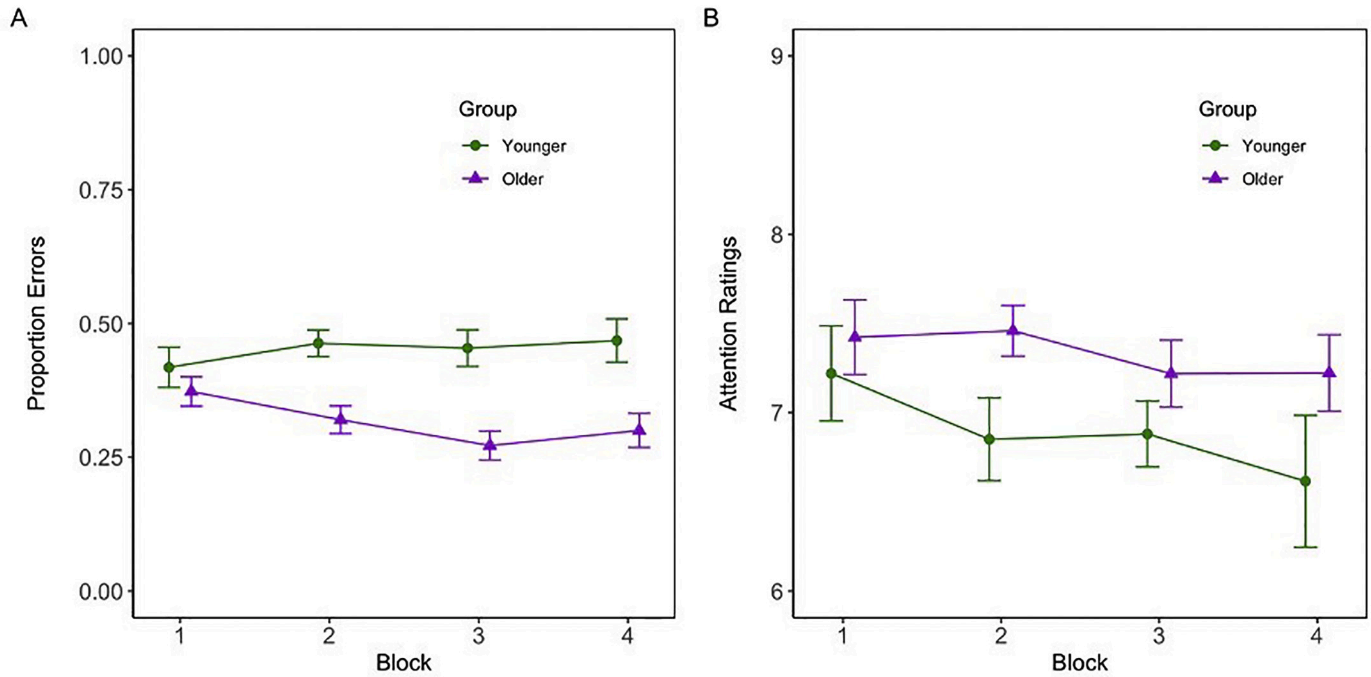

Fig. D1. Visualization of proportion commission errors and mean attention ratings (y-axes) as a function of time on task (x-axis) in the SART. Error bars are the 95% within-subjects' confidence intervals.

Appendix E. Visualization of response frequencies in visual WM task

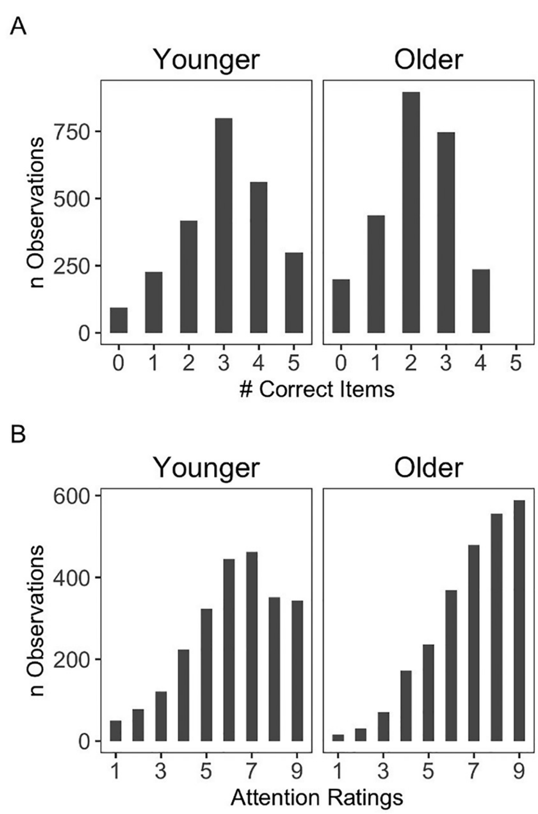

Fig. E1. Number of observations for correctly reported memory items (Panel A) and attention ratings (Panel B) in the visual working memory task. Both panels visualize data from attention-probed trials only (i.e., the data reported in the main analysis). Note that younger adults were presented five and older adults four memory items.

Appendix F. Results of regression analyses on binned attention ratings in the Gibberish detection, reading LTM, and SART

Attention ratings' distribution in the Reading task was skewed towards high responses (see Fig. F1), and binning into three categories led to convergence issues. We repeated the analyses with ratings binned into two categories (ratings 1–7: "low" and 8–9 "high"). First, we computed the trial-by-trial rstanarm model on gibberish detection lapses, including Age Group and Attention Ratings as binary predictors (priors: cauchy(location = 0, scale = 5), $\widehat{R} = 1.02$, effective/total sample size ratio > 0.13, number of effective samples > 542). As reported in Table F1, the results' pattern was identical to that of the main analysis: Higher attention ratings predicted fewer gibberish detection lapses, and older adults' gibberish detection performance was superior to that of younger adults'. Next, we modeled the LTM part of the Reading task (priors: cauchy(location = 0, scale = 5), $\widehat{R} = 1.005$, effective/total sample size ratio > 0.08, number of effective samples > 327). Similar to the main analysis, the model revealed no credible effects of attention or age group on LTM accuracy (see Table F1).

Last, we repeated the analysis of SART commissions. Attention ratings in the SART were skewed towards high responses (see Fig. F1, Panel C), and model with three categories did not converge. Hence, we binned ratings into low (1–7) and high (8–9). The model (priors: cauchy(location = 0, scale = 5), $\widehat{R} = 1.001$, effective/total sample size ratio > 0.35, number of effective samples > 1068) showed that older adults had fewer commission errors, but, contrary to the main analysis, the attention rating term was no longer credible (see Table F1).

In summary, converting attention ratings from a continuous to a categorical predictor led to a more balanced distribution of observations between the factor levels, but the loss of variability resulted in reduction in explanatory power.

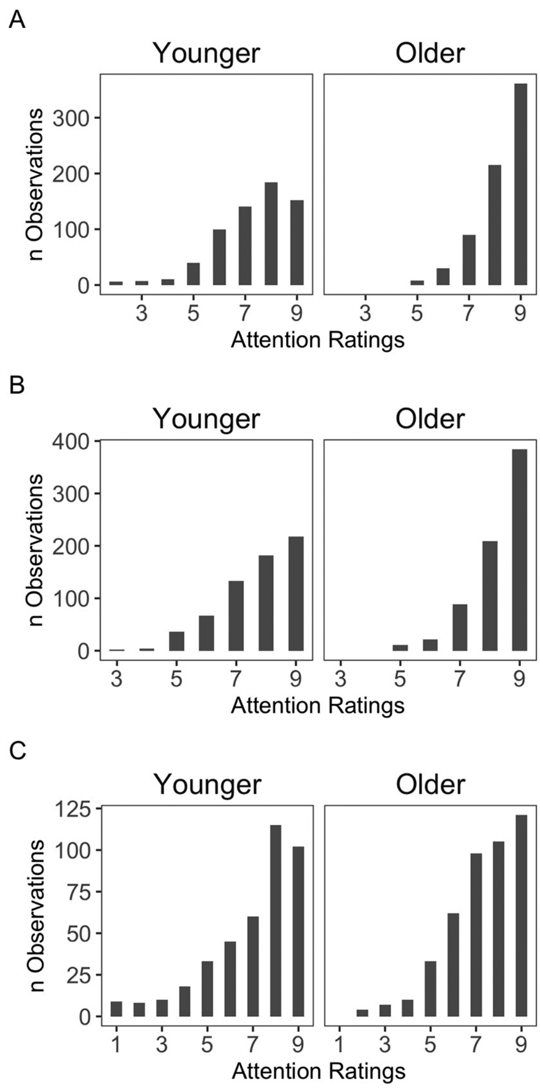

Fig. F1. Number of observations for attention ratings on the gibberish detection (Panel A) and in the long-term memory (Panel B) trials in the reading task. Panel C depicts attention ratings after no-go trials in the SART.

Table F1
Posterior Estimates of the Predictors of the Reading Task and SART, Attention Ratings binned to Low (1–7) and High (8–9).

Parameters	Gibberish detection		LTM Accuracy		SART commissions	
	Mean	95% HDI	Mean	95% HDI	Mean	95% HDI
Intercept	−1.67	[−2.22, 1.19]	1.12	[0.88, 1.39]	−0.15	[−0.56, 0.25]
Attention	**−0.62**	**[−1.19, −0.01]**	−0.26	[−0.55, 0.01]	−0.26	[−0.77, 0.24]
Group	**−0.96**	**[−1.99, −0.009]**	−0.25	[−0.76, 0.24]	**−0.66**	**[−1.24, −0.08]**
Attention × Group	0.44	[−0.59, 1.50]	−0.24	[−0.79, 0.33]	−0.02	[−0.76, 0.71]

Note. Parameter values are on the logit scale. Values in boldface indicate credible effects (i.e., the HDI does not include zero).

References

Adam, K. C. S., & Vogel, E. K. (2017). Confident failures: Lapses of working memory reveal a metacognitive blind spot. *Attention, Perception, & Psychophysics, 79*(5), 1506–1523. https://doi.org/10.3758/s13414-017-1331-8.

Addis, D. R., Musicaro, R., Pan, L., & Schacter, D. L. (2010). Episodic simulation of past and future events in older adults: Evidence from an experimental recombination task. *Psychology and Aging, 25*(2), 369–376. https://doi.org/10.1037/a0017280.

Addis, D. R., Wong, A. T., & Schacter, D. L. (2008). Age-related changes in the episodic simulation of future events. *Psychological Science, 19*(1), 33–41. https://doi.org/10.1111/j.1467-9280.2008.02043.x.

Allen, M., Smallwood, J., Christensen, J., Gramm, D., Rasmussen, B., Gaden Jensen, C., ... Lutz, A. (2013). The balanced mind: The variability of task-unrelated thoughts predicts error-monitoring. *Frontiers in Human Neuroscience, 7*. https://doi.org/10.3389/fnhum.2013.00743.

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

Baumeister, R. F., Hofmann, W., Summerville, A., Reiss, P. T., & Vohs, K. D. (2020). Everyday thoughts in time: Experience sampling studies of mental time travel. *Personality and Social Psychology Bulletin, 0146167220908411*. https://doi.org/10.1177/0146167220908411.

Ben-Shachar, M. S., Makowski, D., & Lüdecke, D. (2020). *Compute and interpret indices of effect size [R package]*. https://github.com/easystats/effectsize.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.

Braver, T. S., & West, R. (2008). Working memory, executive control, and aging. *The Handbook of Aging and Cognition,* 311–372.

Brockmole, J. R., & Logie, R. H. (2013). Age-related change in visual working memory: A study of 55,753 participants aged 8–75. *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00012.

Carriere, J. S. A., Cheyne, J. A., Solman, G. J. F., & Smilek, D. (2010). Age trends for failures of sustained attention. *Psychology and Aging, 25*(3), 569–574. https://doi.org/10.1037/a0019363.

Carstensen, L. L. (1993). Motivation for social contact across the life span: A theory of socio-emotional selectivity. In J. Jacobs (Ed.). *Nebraska Symposium on Motivation. Developmental perspectives on motivation (pp. 209–254)*. University of Nebraska Press.

Carstensen, L. L. (1995). Evidence for a life-span theory of socioemotional selectivity. *Current Directions in Psychological Science, 4*(5), 151–156 JSTOR.

Castel, A. D., Middlebrooks, C. D., & McGillivray, S. (2015). In J. Dunlosky, & S. (Uma) K. Tauber (Vol. Eds.), *Monitoring memory in old age. Vol. 1*Oxford University Presshttps://doi.org/10.1093/oxfordhb/9780199336746.013.3.

Cavanaugh, J. C., & Blanchard-Fields, F. (2018). *Adult development and aging.* Cengage Learning.

Chun, M. M. (2011). Visual working memory as visual attention sustained internally over time. *Neuropsychologia, 49*(6), 1407–1409. https://doi.org/10.1016/j.neuropsychologia.2011.01.029.

Core Team, R. (2018). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Cox, W. M., Klinger, E., & Fadardi, J. S. (Eds.). (2011). Personal aspirations and concerns inventory (PCI). *Handbook of motivational counseling: Goal-based approaches to assessment and intervention with addiction and other problems* (pp. 200–201). (2nd ed.). John Wiley & Sons.

Duarte, A., Hearons, P., Jiang, Y., Delvin, M. C., Newsome, R. N., & Verhaeghen, P. (2013). Retrospective attention enhances visual working memory in the young but not the old: An ERP study. *Psychophysiology, 50*(5), 465–476. https://doi.org/10.1111/psyp.12034.

Einstein, G. O., & Mcdaniel, M. A. (1997). Aging and mind wandering: Reduced inhibition in older adults? *Experimental Aging Research, 23*(4), 343–354. https://doi.org/10.1080/03610739708254035.

Folstein, M. F., Robins, L. N., & Helzer, J. E. (1983). The mini-mental state examination. *Archives of General Psychiatry, 40*(7), https://doi.org/10.1001/archpsyc.1983.01790060110016 812–812.

Fortenbaugh, F. C., DeGutis, J., Germine, L., Wilmer, J. B., Grosso, M., Russo, K., & Esterman, M. (2015). Sustained attention across the life span in a sample of 10,000: Dissociating ability and strategy. *Psychological Science, 26*(9), 1497–1510. https://doi.org/10.1177/0956797615594896.

Frank, D. J., Nara, B., Zavagnin, M., Touron, D. R., & Kane, M. J. (2015). Validating older adults' reports of less mind-wandering: An examination of eye movements and dispositional influences. *Psychology and Aging, 30*(2), 266–278. https://doi.org/10.1037/pag0000031.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Chapman & Hall/CRC.

Giambra, L. M. (1997). Sustained attention and aging: Overcoming the decrement? *Experimental Aging Research, 23*(2), 145–161. https://doi.org/10.1080/03610739708254030.

Goodrich, B., Gabry, J., Ali, I., & Brilleman, S. (2018). *rstanarm: Bayesian applied regression modeling via Stan. (2.17.4.) [computer software]*.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In L. Hasher, & R. T. Zacks (Vol. Eds.), *Psychology of learning and motivation. Vol. 22. Psychology of learning and motivation* (pp. 193–225). Academic Press.

Hertzog, C., & Dunlosky, J. (2011). Metacognition in later adulthood: Spared monitoring can benefit older Adults' self-regulation. *Current Directions in Psychological Science, 20*(3), 167–173. https://doi.org/10.1177/0963721411409026.

Jackson, J. D., & Balota, D. A. (2012). Mind-wandering in younger and older adults: Converging evidence from the sustained attention to response task and reading for comprehension. *Psychology and Aging, 27*(1), 106–119. https://doi.org/10.1037/a0023933.

Jackson, J. D., Weinstein, Y., & Balota, D. A. (2013). Can mind-wandering be timeless? Atemporal focus and aging in mind-wandering paradigms. *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00742.

Jordão, M., Ferreira-Santos, F., Pinho, M. S., & St. Jacques, P. L. (2019). Meta-analysis of aging effects in mind wandering: Methodological and sociodemographic factors. *Psychology and Aging, 34*(4), 531–544. https://doi.org/10.1037/pag0000356.

Kane, M. J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). For whom the mind wanders, and when: An experience-sampling study of working memory and executive control in daily life. *Psychological Science, 18*(7), 614–621.

Kane, M. J., Meier, M. E., Smeekens, B. A., Gross, G. M., Chun, C. A., Silvia, P. J., & Kwapil, T. R. (2016). Individual differences in the executive control of attention, memory, and thought, And Their Associations With Schizotypy. *Journal of Experimental Psychology. General, 145*(8), 1017–1048. https://doi.org/10.1037/xge0000184.

Killingsworth, M. A., & Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science, 330*(6006), https://doi.org/10.1126/science.1192439 932–932.

Klinger, E. (1971). *Structure and functions of fantasy.* Wiley.

Klinger, E. (2009). Daydreaming and fantasizing: Thought flow and motivation. In K. D. Markman, W. M. P. Klein, & J. A. Suhr (Eds.). *Handbook of imagination and mental simulation. (pp. 225–239)*. Psychology Press.

Klinger, E., Barta, S. G., & Maxeiner, M. E. (1980). Motivational correlates of thought content frequency and commitment. *Journal of Personality and Social Psychology, 39*(6), 1222–1237. https://doi.org/10.1037/h0077724.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349.

Krawietz, S. A., Tamplin, A. K., & Radvansky, G. A. (2012). Aging and mind wandering during text comprehension. *Psychology and Aging, 27*(4), 951–958. https://doi.org/10.1037/a0028831.

Krimsky, M., Forster, D. E., Llabre, M. M., & Jha, A. P. (2017). The influence of time on task on mind wandering and visual working memory. *Cognition, 169*, 84–90. https://doi.org/10.1016/j.cognition.2017.08.006.

Kvavilashvili, L., & Rummel, J. (2020). On the nature of everyday prospection: A review and theoretical integration of research on mind-wandering, future thinking, and prospective memory. *Review of General Psychology, 24*(3), 210–237. https://doi.org/10.1177/1089268020918843.

Levinson, D. B., Smallwood, J., & Davidson, R. J. (2012). The persistence of thought: Evidence for a role of working memory in the maintenance of task-unrelated thinking. *Psychological Science, 23*(4), 375–380. https://doi.org/10.1177/0956797611431465.

Lindenberger, U., & Mayr, U. (2014). Cognitive aging: Is there a dark side to environmental support? *Trends in Cognitive Sciences, 18*(1), 7–15. https://doi.org/10.1016/j.tics.2013.10.006.

Loaiza, V. M., & Souza, A. S. (2019). An age-related deficit in preserving the benefits of attention in working memory. *Psychology and Aging, 32*(2), 282.

Logan, G. D. (1985). Executive control of thought and action. *Acta Psychologica, 60*(2), 193–210. https://doi.org/10.1016/0001-6918(85)90055-1.

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory.

Nature Neuroscience, 17(3), 347–356. https://doi.org/10.1038/nn.3655.

Maillet, D., Beaty, R. E., Jordano, M. L., Touron, D. R., Adnan, A., Silvia, P. J., ... Kane, M. J. (2018). Age-related differences in mind-wandering in daily life. *Psychology and Aging, 33*(4), 643–653. https://doi.org/10.1037/pag0000260.

McAvinue, L. P., Habekost, T., Johnson, K. A., Kyllingsbæk, S., Vangkilde, S., Bundesen, C., & Robertson, I. H. (2012). Sustained attention, attentional selectivity, and attentional capacity across the lifespan. *Attention, Perception, & Psychophysics, 74*(8), 1570–1582. https://doi.org/10.3758/s13414-012-0352-6.

McVay, J. C., & Kane, M. J. (2009). Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(1), 196–204. https://doi.org/10.1037/a0014104.

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

McVay, J. C., & Kane, M. J. (2012a). Drifting from slow to “d”oh!’: Working memory capacity and mind wandering predict extreme reaction times and executive control errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(3), 525–549. https://doi.org/10.1037/a0025896.

McVay, J. C., & Kane, M. J. (2012b). Why does working memory capacity predict variation in Reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology. General, 141*(2), 302–320. https://doi.org/10.1037/a0025250.

McVay, J. C., & Kane, M. J. (2013). Dispatching the wandering mind? Toward a laboratory method for cuing “spontaneous” off-task thought. *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00570.

McVay, J. C., Meier, M. E., Touron, D. R., & Kane, M. J. (2013). Aging ebbs the flow of thought: Adult age differences in mind wandering, executive control, and self-evaluation. *Acta Psychologica, 142*(1), 136–147. https://doi.org/10.1016/j.actpsy.2012.11.006.

Mitchell, D. J., Cusack, R., & Cam-CAN (2018). Visual short-term memory through the lifespan: Preserved benefits of context and metacognition. *Psychology and Aging, 33*(5), 841–854. https://doi.org/10.1037/pag0000265.

Mittner, M., Boekel, W., Tucker, A. M., Turner, B. M., Heathcote, A., & Forstmann, B. U. (2014). When the brain takes a break: A model-based analysis of mind wandering. *The Journal of Neuroscience, 34*(49), 16286–16295. https://doi.org/10.1523/JNEUROSCI.2062-14.2014.

Mooneyham, B. W., & Schooler, J. W. (2013). The costs and benefits of mind-wandering: A review. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 67*(1), 11–18. https://doi.org/10.1037/a0031569.

Morey, C. C., & Bieler, M. (2013). Visual short-term memory always requires general attention. *Psychonomic Bulletin & Review, 20*(1), 163–170. https://doi.org/10.3758/s13423-012-0313-z.

Morey, R. D., Rouder, J. N., Jamil, T., Forner, K., & Ly, A. (2018). *Package “bayesfactor”. (0.9.12-4.2) [Computer software]*. https://richarddmorey.github.io/BayesFactor/.

Morrison, A. B., Goolsarran, M., Rogers, S. L., & Jha, A. P. (2014). Taming a wandering attention: Short-form mindfulness training in student cohorts. *Frontiers in Human Neuroscience, 7*. https://doi.org/10.3389/fnhum.2013.00897.

Mrazek, M. D., Smallwood, J., Franklin, M. S., Chin, J. M., Baird, B., & Schooler, J. W. (2012). The role of mind-wandering in measurements of general aptitude. *Journal of Experimental Psychology: General, 141*(4), 788–798. https://doi.org/10.1037/a0027968.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231–259. https://doi.org/10.1037/0033-295X.84.3.231.

Oberauer, K., & Lin, H.-Y. (2017). An interference model of visual working memory. *Psychological Review, 124*(1), 21–59. https://doi.org/10.1037/rev0000044.

Parasuraman, R., Nestor, P. G., & Greenwood, P. (1989). Sustained-attention capacity in young and older adults. *Psychology and Aging, 4*(3), 339–345. https://doi.org/10.1037/0882-7974.4.3.339.

Parks, C. W., Klinger, E., & Perlmutter, M. (1989). Dimensions of thought as a function of age, gender and task difficulty. *Imagination, Cognition and Personality, 8*(1), 49–62. https://doi.org/10.2190/M6GA-J94F-VRV1-77DR.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442.

Plimpton, B., Patel, P., & Kvavilashvili, L. (2015). Role of triggers and dysphoria in mind-wandering about past, present and future: A laboratory study. *Consciousness and Cognition, 33*, 261–276. https://doi.org/10.1016/j.concog.2015.01.014.

Qin, J., Perdoni, C., & He, B. (2011). Dissociation of subjectively reported and behaviorally indexed mind wandering by EEG rhythmic activity. *PLoS One, 6*(9), https://doi.org/10.1371/journal.pone.0023124.

Rey-Mermet, A., & Gade, M. (2018). Inhibition in aging: What is preserved? What declines? A meta-analysis. *Psychonomic Bulletin & Review, 25*(5), 1695–1716. https://doi.org/10.3758/s13423-017-1384-7.

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). Oops!’: Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia, 35*(6), 747–758.

Robison, M. K., & Unsworth, N. (2015). Working memory capacity offers resistance to mind-wandering and external distraction in a context-specific manner. *Applied Cognitive Psychology, 29*(5), 680–690. https://doi.org/10.1002/acp.3150.

Robison, M. K., & Unsworth, N. (2017). Working memory capacity and mind-wandering during low-demand cognitive tasks. *Consciousness and Cognition, 52*, 47–54. https://doi.org/10.1016/j.concog.2017.04.012.

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

Ruby, F. J. M., Smallwood, J., Engen, H., & Singer, T. (2013). How self-generated thought shapes mood—The relation between mind-wandering and mood depends on the socio-temporal content of thoughts. *PLoS One, 8*(10), Article e77554. https://doi.org/10.1371/journal.pone.0077554.

Rummel, J., & Boywitt, C. D. (2014). Controlling the stream of thought: Working memory capacity predicts adjustment of mind-wandering to situational demands. *Psychonomic Bulletin & Review, 21*(5), 1309–1315. https://doi.org/10.3758/s13423-013-0580-3.

Schooler, J. W., Smallwood, J., Christoff, K., Handy, T. C., Reichle, E. D., & Sayette, M. A. (2011). Meta-awareness, perceptual decoupling and the wandering mind. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2011.05.006.

Schooler, J. W. (2002). Re-representing consciousness: Dissociations between experience and meta-consciousness. *Trends in Cognitive Sciences, 6*(8), 339–344.

Schubert, A.-L., Frischkorn, G. T., & Rummel, J. (2019). The validity of the online thought-probing procedure of mind wandering is not threatened by variations of probe rate and probe framing. *Psychological Research Psychologische Forschung*. https://doi.org/10.1007/s00426-019-01194-2.

Seli, P., Beaty, R. E., Cheyne, J. A., Smilek, D., Oakman, J., & Schacter, D. L. (2018). How pervasive is mind wandering, really? *Consciousness and Cognition, 66*, 74–78. https://doi.org/10.1016/j.concog.2018.10.002.

Seli, P., Beaty, R. E., Cheyne, J. A., Smilek, D., & Schacter, D. L. (2018). *How pervasive is mind wandering, really?* .

Seli, P., Cheyne, J. A., & Smilek, D. (2012). Attention failures versus misplaced diligence: Separating attention lapses from speed–accuracy trade-offs. *Consciousness and Cognition, 21*(1), 277–291. https://doi.org/10.1016/j.concog.2011.09.017.

Seli, P., Jonker, T. R., Cheyne, J. A., Cortes, K., & Smilek, D. (2015). Can research participants comment authoritatively on the validity of their self-reports of mind wandering and task engagement? *Journal of Experimental Psychology: Human Perception and Performance, 41*(3), 703–709. https://doi.org/10.1037/xhp0000029.

Seli, P., Kane, M. J., Metzinger, T., Smallwood, J., Schacter, D. L., Maillet, D., ... Smilek, D. (2018). The family-resemblances framework for mind-wandering remains well clad. *Trends in Cognitive Sciences, 22*(11), 959–961. https://doi.org/10.1016/j.tics.2018.07.007.

Seli, P., Maillet, D., Smilek, D., Oakman, J. M., & Schacter, D. L. (2017). Cognitive aging and the distinction between intentional and unintentional mind wandering. *Psychology and Aging, 32*(4), 315–324. https://doi.org/10.1037/pag0000172.

Seli, P., Smilek, D., Ralph, B. C. W., & Schacter, D. L. (2018). The awakening of the attention: Evidence for a link between the monitoring of mind wandering and prospective goals. *Journal of Experimental Psychology: General, 147*(3), 431–443. https://doi.org/10.1037/xge0000385.

Smallwood, J. (2013). Distinguishing how from why the mind wanders: A process–occurrence framework for self-generated mental activity. *Psychological Bulletin, 139*(3), 519–535. https://doi.org/10.1037/a0030010.

Smallwood, J., Riby, L., Heim, D., & Davies, J. B. (2006). Encoding during the attentional lapse: Accuracy of encoding during the semantic sustained attention to response task. *Consciousness and Cognition, 15*(1), 218–231. https://doi.org/10.1016/j.concog.2005.03.003.

Smallwood, J., Ruby, F. J. M., & Singer, T. (2013). Letting go of the present: Mind-wandering is associated with reduced delay discounting. *Consciousness and Cognition, 22*(1), 1–7. https://doi.org/10.1016/j.concog.2012.10.007.

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

Song, X., & Wang, X. (2012). Mind wandering in Chinese daily lives – An experience sampling study. *PLoS One, 7*(9), Article e44423. https://doi.org/10.1371/journal.pone.0044423.

Staub, B., Doignon-Camus, N., Bacon, É., & Bonnefond, A. (2014). The effects of aging on sustained attention ability: An ERP study. *Psychology and Aging, 29*(3), 684–695. https://doi.org/10.1037/a0037067.

Staub, B., Doignon-Camus, N., Marques-Carneiro, J. E., Bacon, É., & Bonnefond, A. (2015). Age-related differences in the use of automatic and controlled processes in a situation of sustained attention. *Neuropsychologia, 75*, 607–616. https://doi.org/10.1016/j.neuropsychologia.2015.07.021.

Stawarczyk, D. (2018). Phenomenological properties of mind-wandering and day-dreaming: A historical overview and functional. In K. C. R. Fox, & K. Christoff (Eds.). *The Oxford handbook of spontaneous thought: Mind-wandering, creativity, and dreaming (p. 193)*. Oxford University Press.

Stawarczyk, D., Cassol, H., & D’Argembeau, A. (2013). Phenomenology of future-oriented mind-wandering episodes. *Frontiers in Psychology,* 4. https://doi.org/10.3389/fpsyg.2013.00425.

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

Steindorf, L., & Rummel, J. (2017). “I should not forget the apples!”—Mind-wandering episodes used as opportunities for rehearsal in an interrupted recall paradigm. *Applied Cognitive Psychology, 31*(4), 424–430. https://doi.org/10.1002/acp.3328.

Strunk, J., Morgan, L., Reaves, S., Verhaeghen, P., & Duarte, A. (2018). Retrospective attention in short-term memory has a lasting effect on long-term memory across age. *The Journals of Gerontology: Series B*. https://doi.org/10.1093/geronb/gby045.

Tomporowski, P. D., & Tinsley, V. F. (1996). Effects of memory demand and motivation on sustained attention in young and older adults. *The American Journal of Psychology, 109*(2), 187–204. JSTOR https://doi.org/10.2307/1423272.

Unsworth, N., & McMillan, B. D. (2013). Mind wandering and reading comprehension: Examining the roles of working memory capacity, interest, motivation, and topic experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(3), 832–842. https://doi.org/10.1037/a0029669.

Unsworth, N., & McMillan, B. D. (2014a). Trial-to-trial fluctuations in attentional state

and their relation to intelligence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(3), 882–891. https://doi.org/10.1037/a0035544.
Unsworth, N., & McMillan, B. D. (2014b). Similarities and differences between mind-wandering and external distraction: A latent variable analysis of lapses of attention and their relation to cognitive abilities. *Acta Psychologica, 150*, 14–25. https://doi.org/10.1016/j.actpsy.2014.04.001.
Unsworth, N., & Robison, M. K. (2016). The influence of lapses of attention on working memory capacity. *Memory & Cognition, 44*(2), 188–196. https://doi.org/10.3758/s13421-015-0560-0.
Unsworth, N., & Robison, M. K. (2017). The importance of arousal for variation in working memory capacity and attention control: A latent variable pupillometry study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(12), 1962–1987. https://doi.org/10.1037/xlm0000421.
Verhaeghen, P. (2011). Aging and executive control: Reports of a demise greatly exaggerated. *Current Directions in Psychological Science, 20*(3), 174–180. https://doi.org/10.1177/0963721411408772.
Wagenmakers, E., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology, 100*(3), 426–432. https://doi.org/10.1037/a0022790.
Warden, E. A., Plimpton, B., & Kvavilashvili, L. (2019). Absence of age effects on spontaneous past and future thinking in daily life. *Psychological Research, 83*(4), 727–746. https://doi.org/10.1007/s00426-018-1103-7.
Weinstein, Y. (2017). Mind-wandering, how do I measure thee with probes? Let me count the ways. *Behavior Research Methods,* 1–20. https://doi.org/10.3758/s13428-017-0891-9.
Weinstein, Y., De Lima, H. J., & van der Zee, T. (2017). Are you mind-wandering, or is your mind on task? The effect of probe framing on mind-wandering reports. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-017-1322-8.
Wiemers, E. A., & Redick, T. S. (2019). The influence of thought probes on performance: Does the mind wander more if you ask it? *Psychonomic Bulletin & Review, 26*(1), 367–373. https://doi.org/10.3758/s13423-018-1529-3.
Zavagnin, M., Borella, E., & De Beni, R. (2014). When the mind wanders: Age-related differences between young and older adults. *Acta Psychologica, 145*, 54–64. https://doi.org/10.1016/j.actpsy.2013.10.016.
Zedelius, C. M., Broadway, J. M., & Schooler, J. W. (2015). Motivating meta-awareness of mind wandering: A way to catch the mind in flight? *Consciousness and Cognition, 36*, 44–53. https://doi.org/10.1016/j.concog.2015.05.016.