Journal of Memory and Language 121 (2021) 104276

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Attending to encode: The role of consistency and intensity of attention in learning ability

Ashley L. Miller *, Nash Unsworth

Department of Psychology, University of Oregon, United States

A R T I C L E I N F O

Keywords:
Attention
Attentional lapses
Learning
Individual differences
Pupillometry

A B S T R A C T

The present study examined how variation in the amount of attention devoted to items (intensity) and the consistency with which attention is maintained on task (consistency) are related to each other and to overall learning abilities. In two experiments, participants completed measures of working memory (WM), long-term memory (LTM), motivation, and a paired associates (PA) cued recall task with thought probes embedded throughout the encoding phase of each word-pair list. In Experiment 2, pupil diameter was also simultaneously recorded during encoding of the PA task to provide an index of the intensity of attention. Results collectively suggested that the most successful learners were those who were both less susceptible to lapses of attention (high consistency) and had larger pupil dilation at encoding (high intensity). Critically, while attentional lapses and pupil dilation were negatively related to one another—both between and within subjects—each aspect of attention accounted for unique variance in associative learning even after accounting for WM, LTM, and motivation. Follow-up analyses further revealed that, while intensity and consistency were both related to motivation and (to a lesser extent) general LTM abilities, motivation was a greater determinant of the consistency of attention. Therefore, it appears that the intensity and consistency of attention are likely distinct, multifaceted constructs that are differentially influenced by a variety of factors and play an important role in learning.

Introduction

Attention abilities are important for performance on a variety of cognitive tasks (e.g., Engle & Kane, 2004; Kane & Engle, 2002; Redick et al., 2016; Unsworth & McMillan, 2014). Work from our laboratory (e.g., Unsworth & Robison, 2015; 2017a; 2017b) has specifically focused on two aspects of these attention abilities: the intensity of attention (the amount of attention/attentional effort devoted to items) and the consistency of attention (how consistently individuals are able to keep their attention on task rather than off task). This work has shown that intensity and consistency are important predictors of working memory and attention control abilities. Recently, we (Miller, Gross, & Unsworth, 2019; Miller & Unsworth, 2020) extended these results to the realm of learning and memory. For example, using pupillary responses as an index of the intensity of attention, we (Miller & Unsworth, 2020) demonstrated that the best learners on an associative memory task devoted more attention to items at encoding than those who worst learned the task. Others have similarly suggested that the best learners are also better able to consistently maintain attention on task. Namely, the best learners tend to be the least susceptible to off task thoughts/lapses of attention (e.g., mind wandering) during encoding (e.g., Seibert & Ellis, 1991; Xu & Metcalfe, 2016). However, existing research has yet to account for the possible role of theoretically important third variables (e.g., motivation), and no study has simultaneously examined both aspects of attention. Thus, it is not only unclear whether factors like motivation are responsible for effects previously attributed to intensity and consistency, but it is also unclear to which extent intensity and consistency reflect similar or distinct processes. The present study adopted an individual differences approach to investigate these questions with the ultimate goal of better elucidating the more nuanced features of successful learning abilities.

Background

Encoding information into long-term memory is a prime example of an attention demanding process that requires substantial effort. Research consistent with this notion demonstrates that under conditions of divided attention, items are weakly encoded, and chances of retrieval

* Corresponding author at: Department of Psychology, University of Oregon, 1227 University St, Eugene, OR 97403, United States.
E-mail address: amiller8@uoregon.edu (A.L. Miller).

https://doi.org/10.1016/j.jml.2021.104276
Received 16 December 2019; Received in revised form 7 June 2021; Accepted 26 June 2021

are lower than when attention is fully devoted to to-be-remembered material (Anderson, Craik, & Naveh-Benjamin, 1998; Baddeley, Lewis, Eldridge, & Thomson, 1984). However, in typical behavioral experiments it is not always possible to directly track how attention and effort are allocated to items during learning. To circumvent this issue, we (Miller et al., 2019; Miller & Unsworth, 2020; Unsworth & Miller, 2021; Unsworth & Robison, 2015; 2017a; 2017b; 2020; Unsworth, Miller, & Robison, 2020) and others (e.g., Ariel & Castel, 2014; Beatty & Lucero-Wagoner, 2000; Just & Carpenter, 1993; Kahneman, 1973; Kahneman & Beatty, 1966; Papesh, Goldinger, & Hout, 2012; Porter, Troscianko, & Gilchrist, 2007) have relied on pupil dilation as an online indicator of intensive attention allocation. For example, Hess and Polt (1964) had participants complete a series of multiplication problems while the diameter of their pupils was recorded. Results revealed more difficult math problems were associated with larger pupil dilations, presumably because these problems were the most attentionally demanding. Pupils also dilate as a function of the number of items being maintained in working memory (see Kahneman & Beatty, 1966; Peavler, 1974; Unsworth & Robison, 2015). Results such as these led Kahneman (1973) to claim that task evoked pupillary responses (i.e., TEPRs)—changes in pupil dilation relative to baseline levels—correspond to the intensive aspect of attention and are a reliable psychophysiological indicator of the amount of attentional effort devoted to a given item (i.e., the "intensity of attention"; see also Just & Carpenter, 1993).

There is a lengthy history of using pupillary responses to examine intensive attention allocation at encoding (e.g., Beatty & Lucero-Wagoner, 2000; Engle, 1975; Goldinger & Papesh, 2012; Heaver & Hutton, 2011; Janisse, 1977; Kafkas & Montaldi, 2011; Papesh et al., 2012). For instance, Ariel and Castel (2014) demonstrated that items high in value (e.g., words worth 10 points) were associated with larger pupillary responses at encoding and better recall than items low in value (e.g., words worth 1 point). Increased pupil dilation and better memory performance has also been observed for deeply encoded items compared to shallowly encoded items (Taikh, 2014; see Otero, Weekes, & Hutton, 2011 for similar results at retrieval), as well as when participants report using effective encoding strategies (e.g., mental imagery, sentence generation) compared to ineffective encoding strategies (e.g., rehearsal; Miller & Unsworth, 2020). Thus increased depth of processing and more effective encoding strategies seem to require the implementation of more attentional effort (i.e., more attentional resources; Craik & Byrd, 1982). Taken altogether, these results suggest that items that receive a greater intensity of attention at encoding, as indexed by pupil dilation, tend to be recalled best.

The results above provide important evidence for the promise of using pupillary responses to track the intensity of attention at encoding, but we do not mean to suggest that increased pupil dilation will always correspond with better subsequent memory. We suggest that the relations are more nuanced than that. As noted by Kanfer (1987; see also Locke & Latham, 1990), the relationship between effort and performance is not always necessarily positive and linear. For example, an individual with inadequate cognitive ability may see little to no gains in performance despite their best efforts (e.g., Kanfer & Ackerman, 1989), akin to laboring in vein. Considering that TEPRs are sensitive to mental effort and arousal more generally (Granholm & Steinhauer, 2004), we expect similar situations to arise in which larger pupil dilation at encoding will not always be associated with enhanced recall, such as when individuals report using ineffective encoding strategies (see Miller & Unsworth, 2020). Note other research using item recognition paradigms has even shown that *decreased* pupil dilation, or pupil constriction, is associated with better subsequent memory (Gross & Dobbins, 2021; Kafkas & Montaldi, 2011; Wetzel, Einhäuser, & Widmann, 2020). A recent study (Kafkas, 2021) proposed that these inconsistent findings regarding the direction of the pupillary subsequent memory effect is at least partially attributable to the moderating influence of stimulus novelty. All in all, though, because pupillary subsequent memory effects are inherently correlational in nature, there are likely a number of factors that influence whether they will be found.

Nonetheless, when studies use paired associates (PA) paradigms—as opposed to item recognition paradigms—and directly manipulate item-level effects in the absence of other confounding manipulations, the results largely suggest that items associated with larger pupillary responses at encoding are remembered better because these items receive greater intensity of attention. Critically, though, recent work from our laboratory has found similar and consistent findings at the between-subjects level. Miller and Unsworth (2020) used a PA cued recall task in which participants learned pairs of words as pupillary responses were simultaneously monitored. Results revealed remarkably stable correlations between TEPRs and associative learning ability (i.e., PA recall accuracy) across two experiments (Experiment 1 $r = .35$; Experiment 2 $r = .37$; N's over 120 in each case). The best learners showed a large ramp up in pupil dilation during the encoding period, whereas the worst learners showed no such increase. These TEPRs continued to explain unique variance in recall accuracy even when controlling for the influence of strategy use, working memory (WM), and long-term memory (LTM) ability. Thus, substantial individual differences exist in the intensity of attention, which is an important predictor of successful learning and memory performance in PA paradigms.

While the *intensity* of attention refers to the amount of attention allocated to a task, the *consistency* of attention refers to how regularly (on an item-by-item basis) an individual allocates attention to on task processing. Most attention will generally be devoted to performing the task itself, but attention can also be allocated to a variety of other processes during a given task (e.g., Kanfer & Ackerman, 1989). For instance, attention may be allocated toward internal, task-unrelated thoughts and ruminations (mind-wandering) or toward external stimuli unrelated to the task at hand (external distraction). External distraction and mind-wandering, as well as the related phenomenon of mind-blanking (i.e., episodes of zoning out/absence of thought believed to reflect more extreme forms of task disengagement; Ward & Wegner, 2013), can altogether be characterized as attentional lapses or, more precisely, off task thoughts. Critically, research largely suggests that task performance is worse when participants report an attentional lapse on the preceding trial (McVay & Kane, 2010a; Unsworth & McMillan, 2014); attention becomes "decoupled" from external stimuli during these episodes (Smallwood, Baracaia, Lowe, & Obonsawin, 2003; Smallwood & Schooler, 2006).

A number of laboratory techniques have been developed to examine one's ability to consistently maintain attention on task, and, more specifically, one's ability to prevent recurrent lapses of attention. However, we (e.g., Robison, Miller, & Unsworth, 2020; Unsworth & McMillan, 2014; Unsworth & Robison, 2017b, 2018) and others (e.g., Kane et al., 2007; McVay & Kane, 2009, 2012; Seli, Cheyne, Xu, Purdon, & Smilek, 2015; Smallwood et al., 2003; Smallwood & Schooler, 2006; Stawarczyk, Majerus, Maj, Van der Linden, & D'Argembeau, 2011) have relied upon the thought-probe technique. The thought-probe technique involves periodically stopping participants throughout a task and directly asking them to report whether their attention, immediately prior to the appearance of the probe, was focused on task or off task (mind-wandering, externally distracted, or mind-blanking). When examining tasks that place increased demands on attention, research has typically found that individuals who report a greater proportion of off task thoughts (less consistency) do not perform as well as individuals who can consistently keep their attention on task (e.g., Kane et al., 2016, 2017; McVay & Kane, 2010a, 2012; Robison & Unsworth, 2015; Unsworth et al., 2020; Unsworth et al., in press). Proportions of off task thought have also been shown to relate to other common indicators of consistency, such as reaction time (RT) coefficient of variation (Kane et al., 2016; Unsworth, 2015), periodic performance failures, and variation in baseline pupil diameter (Unsworth and Robison, 2017b). In sum, the thought probe technique is a reliable, valid indicator of one's ability to consistently keep attention on task in a variety of laboratory and everyday settings (e.g., Kane et al., 2007, 2017; McVay, Kane, &

Kwapil, 2009; Unsworth & McMillan, 2017; Unsworth, McMillan, Brewer, & Spillers, 2012).

Of particular relevance to the present study, research has begun to use thought-probes in LTM tasks. This work suggests that lapses of attention frequently occur during learning and are associated with poorer subsequent memory performance (Garlitch & Wahlheim, 2020; Maillet & Rajah, 2014; Metcalfe & Xu, 2016; Smallwood, Baracaia, Lowe, & Obonsawin, 2003; Thomson, Smilek, & Besner, 2014; Xu & Metcalfe, 2016). Susceptibility to attentional lapses (inconsistency) has also been associated with individual differences in learning and memory abilities. Namely, Seibert and Ellis (1991) revealed that greater proportions of off task thoughts were largely associated with impaired recall (Experiment 1 $r(44) = -.72$; Experiment 2 $r(44) = -.67$), meaning individuals who are better able to consistently keep their attention on task tend to display superior memory performance relative to individuals who are less able to do so. More recently, Xu and Metcalfe (2016) replicated these results in two of their three experiments that examined the relation between mind-wandering and associative learning, Experiment 2 $r(23) = -.46$; Experiment 3 $r(84) = -.22$ (see also Garlitch & Wahlheim, 2020; Metcalfe & Xu, 2016). Overall, it is apparent that the consistency of attention (as indexed by proportions of off task thought/attentional lapses) is another important factor necessary for successful learning.

Present study

The research reviewed thus far suggests that two aspects of attention are important for learning and memory performance: (1) how consistently individuals are able to keep their attention on task rather than off task during learning (consistency), and (2) the amount of attention/attentional effort devoted to learning (intensity). However, individual differences research on these processes is still in its infancy. Most existing research has failed to account for the possible roles of theoretically meaningful third variables, which is problematic because focusing on a single variable to predict learning results in a narrowly defined learning construct that neglects many of its more nuanced features. That is, most constructs have several underlying influences (some of which may account for variance explained by the single construct in question). This is why regression analyses consisting of multiple predictors are critical. Otherwise—if only a single predictor is under consideration—there is no way to determine whether learning-related differences in one construct are distinct from learning-related differences in another construct. As such, the present study had four broad aims.

First, we sought to better understand why the ability to consistently maintain attention on task during learning predicts associative learning ability. Prior work suggests that one variable worthy of consideration is motivation, insofar that those who are the most motivated to perform well on a given task tend to experience a lower proportion of off task thoughts (more consistency) during said task (Robison & Unsworth, 2018; Seli et al., 2015; Unsworth & McMillan, 2013). Theoretically, in an attempt to maximize performance, individuals who are more motivated to perform well should more consistently direct attention to the task at hand. In being more consistently focused on the task (i.e., less susceptible to attentional capture by external or internal sources across trials), these high motivation individuals should perform better on the task. Note that as time-on task increases and it becomes harder to maintain attention across trials, attentional lapses become more difficult to inhibit leading to an increase in lapses and subsequent decreases in performance (see Thomson et al., 2014). Those who are the most motivated to perform well may also be more likely to continue in their attempts to sustain attention across the entire task, resulting in fewer attentional lapses (better consistency) and better performance. Those who are the least motivated to perform well, on the other hand, may eventually disengage from the task at hand in favor of entertaining task-unrelated thoughts, resulting in worse performance. For example, a person with low motivation may direct their attention towards intentional forms of mind-wandering (e.g., readily thinking about more pleasant things like a post-COVID vacation; see Seli, Schacter, Risko, & Smilek, 2019).

While motivation is likely one variable that may explain why lapses of attention (inconsistency) predict learning and memory failures, prior research suggests that executive attention abilities might be important too. Namely, working memory (WM) abilities tend to predict lapses of attention/off task thoughts at the latent level (e.g., Unsworth & Robison, 2017b), with latent correlations typically around $r = -.20$ but zero-order correlations far smaller (Kane et al., 2016; McVay & Kane, 2012; Robison et al., 2020; Unsworth et al., in press). Nonetheless, the executive-attention view of WM (see Engle & Kane, 2004) theorizes that the ability to control attention in the presence of interference or other potent distractors is a critical mechanism responsible for WM's relation with other complex cognitive tasks (such as fluid reasoning, reading comprehension, and learning). This view predicts that high WM individuals should display better task performance partly because they are better able to consistently maintain their attention on task and experience fewer off task thoughts (see also McVay & Kane, 2010b). As such, it seems possible that prior results observing a negative correlation between off task thoughts (inconsistency) and learning may be explained by either motivation or executive attention abilities like WM.

Measures of WM may also serve as a proxy of one's overall attentional resource capacity (Conway & Engle, 1994; Engle & Kane, 2004), which is especially relevant to the second aim of the present study: to better understand why the best learners also tend to allocate more attention (intensity) to items at study. Attention allocation models have long suggested that individuals differ in the amount of available attentional resources at their disposal (e.g., Kanfer & Ackerman, 1989). Intensity should, therefore, have a ceiling, such that one's attentional resource capacity should place upper bounds on the amount of attention that can be allocated to a task at any given moment. An additional key aspect of attention allocation models is the notion that people rarely use all of their available attentional resources at any time (Hockey, 1997; Kalsbeek, 1968; Kanfer & Ackerman, 1989; Schmidtke, 1976). Rather, people seem to allocate an initial proportion of their attention to a given task, with some attention being spared (Ackerman, 2011; Hockey, 1997, 2013; Kalsbeek, 1968; Kanfer & Ackerman, 1989; Schmidtke, 1976). Put differently, individuals are inclined to conserve some resources for later use.[1]

One variable that may impact one's tendency to conserve attentional resources is motivation. Namely, motivation should influence how much of one's available capacity is allocated to a task, insofar that those who are more motivated to perform well should allocate a greater proportion of their available resources (more intensity) to learning. This is to say that those who have the most available resources and are also the most motivated to expend those resources should have the largest intensity. Importantly, though, motivation and the intensity of attention should contribute both shared and unique variance when predicting learning abilities. By way of illustration, two individuals may allocate the same amount of attention (intensity) to a given item but for very different reasons. One individual may have fewer available resources but be sufficiently motivated to perform well and thereby allocate the majority of their available resources to learning. On the other hand, another

[1] Attentional resource capacity may also influence consistency, but we suspect such a relation would primarily be attributed to shared variance with intensity. If two individuals who differed only in attentional resource capacity allocated all of their resources to a task, the individual with fewer resources would presumably have lower intensity than would the individual with more resources. In having lowered intensity, the individual with fewer resources would be in a lowered state of task readiness/task engagement on every trial. This lowered state of task engagement would then result in an increased likelihood of experiencing an attentional lapse, as potent task-irrelevant concerns would be more likely to break into the focus of attention.

individual with more available resources may be unmotivated to perform well and may consequently allocate a smaller proportion of their available resources to learning. Thus, by not being motivated, the costs of allocating effort seemingly outweigh the benefits of allocating effort (Shenhav et al., 2013), meaning those with low motivation should allocate less attention to the task at hand by conserving more of their available attentional resources for later use. Such an example therefore illustrates how motivational processes can determine how engaged an individual is with learning. Since prior research examining the role of the intensity of attention has yet to directly account for motivation and has found inconsistent relations with WM (see Miller et al., 2019; Miller & Unsworth, 2020), the present study sought to clarify whether these variables explain the relationship between the intensity of attention and learning ability.

Considering that the existing literature has yet to simultaneously examine both consistency and intensity, the third aim of the present study was to determine whether these two aspects of attention reflect different, unique abilities or whether they reflect the same, general ability to control aspects of one's attention at learning. Theoretically, intensity and consistency should be related to some degree. When experiencing a lapse of attention, attention seemingly shifts towards whatever is causing the lapse (e.g., internally in the case of mind-wandering or externally in the case of external distraction), meaning less attention (less intensity) is directed to the current target stimulus. In support of this notion, Unsworth et al. (2020) showed that when participants reported experiencing an attentional lapse, pupillary responses were smaller than when participants reported being fully on task (see also Hutchison et al., 2020). In the case of learning, this means that a lapse of attention (inconsistency) should be associated with temporarily reduced intensity, resulting in a weaker memory representation that is less likely to be remembered. In terms of individual differences, an individual who is less engaged with the task (due to low intensity) might also experience more lapses of attention (less consistency) as potent task-irrelevant concerns would be more likely to break into the focus of attention. Thus, prior results indicating the intensity of attention is an important predictor of successful learning (Miller et al., 2019; Miller & Unsworth, 2020) may actually be explained by consistency (or vice-versa).

Of course, despite the expected link between intensity and consistency, each may still reflect distinct aspects of attention. Presumably, the best performers should be high in both intensity and consistency, whereas the worst performers should be low in both intensity and consistency. However, it seems possible that there are individuals who are high in one of these aspects and low in the other. For example, some individuals may devote a lot of attention to items but struggle to consistently stay on task (high intensity-low consistency). These individuals would demonstrate high levels of performance when they are on task, but they would also have various trials with low performance due to recurrent attentional lapses. Other individuals may allocate low levels of attention to the current task (perhaps due to decreased attentional resource capacity), leading to lowered levels of task performance. But their allocation of attention may not change much from trial-to-trial (low intensity-high consistency), possibly because they are highly motivated. Since unique effects in simultaneous regression analyses represent the influence of a variable that is statistically independent of the other variables included in the model, a critical point of analysis was to therefore determine whether intensity and consistency uniquely predict learning ability when taking each other (and other important variables) into account. If intensity and consistency are unique predictors of learning, the final aim of the present study was to identify factors that differentially influence each aspect of attention (e.g., motivation and WM abilities).

The current study provides the first critical test of the ideas described above, whereby we measure both consistency and intensity as well as factors suspected of influencing these two aspects of attention. Specifically, Experiment 1 required participants to complete a PA cued recall task with thought probes embedded during the encoding phase of each word-pair list. PA recall accuracy served as our measure of associative learning ability, whereas proportions of off task thought indexed the consistency of attention. Measures of motivation, WM, and general long-term memory (LTM) abilities were also administered. Measures of general LTM abilities were included in an attempt to control for the influence of broad episodic memory abilities independent of associative learning, because (1) substantial and robust individual differences exist in these abilities (Unsworth, 2019) and (2) general LTM abilities explain substantial variance in associative learning (Miller & Unsworth, 2020). We reasoned that, in order to be considered important and unique predictors of associative learning ability, consistency and intensity would need to account for variance over and beyond that accounted for by other meaningful variables, including general LTM abilities. Experiment 2 adopted the same procedure as Experiment 1, but pupil dilation was simultaneously recorded during the PA cued recall task to provide an index of the intensity of attention. Assessing individual differences in various aspects of attention during learning in conjunction with other important variables will allow us to better understand the complex nature of how attention supports successful learning.

Experiment 1

Experiment 1 was conducted to accomplish two goals. First, before examining any pupillary effects, we sought to replicate prior work demonstrating that items studied during an attentional lapse are less likely to be remembered compared to when attention is fully focused on the task (Garlitch & Wahlheim, 2020; Maillet & Rajah, 2013, 2014; Metcalfe & Xu, 2016; Seibert & Ellis, 1991; Smallwood, Baracaia, Lowe, & Obonsawin, 2003; Thomson, Smilek, & Besner, 2014; Xu & Metcalfe, 2016), and people who report fewer attentional lapses during encoding (more consistency) display better learning ability (meta-analytic $r = -.44$, $SE = .11$, 95% CI [−.66, −.21], $N = 513$; Garlitch & Wahlheim, 2020; Metcalfe & Xu, 2016; Seibert & Ellis, 1991; Thomson, Smilek, & Besner, 2014; Xu & Metcalfe, 2016). The second, primary goal of Experiment 1 was to examine whether the consistency of attention (as indexed by proportions of off task thought) still predicts learning ability (as indexed by overall PA recall accuracy) when taking other important variables into account. As discussed previously, prior results claiming that failures in learning are the result of attentional lapses (inconsistency) may either overestimate the effect of consistency on learning or may even be explained by other variables. For instance, inconsistency could arise due to low motivation whereby those who are less motivated to perform well consistently disengage from the task in favor of entertaining more off task thoughts (Robison & Unsworth, 2018; Seli et al., 2015; Unsworth & McMillan, 2013). Such an account would not only imply that there should be a negative correlation between motivation and proportions of off task thought, but also that motivation should explain the relationship between off task thought and PA recall accuracy.

Variation in consistency could also be attributed to individual differences in WM (Kane et al., 2016; McVay & Kane, 2009, 2012, Robison & Unsworth, 2015, 2018, Unsworth & McMillan, 2013, 2014, 2017; Unsworth & Robison, 2017b), insofar that individuals with superior WM abilities could be better at actively maintaining goal relevant information while preventing irrelevant information (e.g., personal concerns) from capturing attention away from the task at hand (McVay & Kane, 2010a, 2010b). According to this view, WM should negatively correlate with proportions of off task thought (inconsistency) and also explain the relationship between off task thought and PA recall accuracy (learning ability). Consequently, the primary analysis of interest was a simultaneous regression with motivation and WM included as predictors (in addition to more general LTM abilities). If the consistency of attention continues to uniquely predict associative learning when controlling for these factors, then we can be more confident in the notion that the consistency of attention is an important explanatory factor of successful learning.

Method

Participants and procedure

A total of 148 participants (60% female) were recruited from the human subject pool at the University of Oregon. Two participants were excluded for being over the age of 35. All other participants were between the ages of 18 and 34 ($M = 19.67$, $SD = 2.03$) and were proficient English speakers. All eligible participants were awarded course research credit for participation. After obtaining informed consent and demographic information, all participants completed three measures of WM: the operation span task (Ospan), the symmetry span task (Symspan), and the reading span task (Rspan). Upon completion of the WM tasks, participants were then administered a paired associates (PA) cued recall task with thought probes embedded throughout the encoding phase of each list. The PA task ended with a questionnaire asking participants to report their motivation to do well on the task. Next, participants completed a delayed free recall task, followed by a picture source recognition task. Two participants were excluded due to being outliers on the PicSource (i.e., recall accuracy was below 10%), and 5 additional participants were excluded for not recalling a single word on any of the PA cued recall lists (final $N = 139$). Of note, participants completed the tasks reported herein as part of a larger experimental test battery lasting approximately 2 h. Since the other tasks administered during the experimental session do not relate to the current study, they are not reported.

WM tasks

Ospan. Participants solved a series of elementary math problems while remembering unrelated letters. First, on computer participants were presented with a math operation (e.g., $(4 \times 1) + 2 = ?$) in which they had to click the mouse to indicate that they had solved the problem. A new screen then appeared with an answer to the math solution (e.g., 6), whereby participants had to indicate if the answer listed onscreen was correct of incorrect via mouse click (e.g., in the case above, the answer 6 would be correct). Upon completion of the math operation, participants were then presented with a letter (i.e., F, H, J, K, L, N, P, Q, R, S, T, and Y) for 1 s. Immediately following letter presentation, the next math problem was presented. Set sizes varied randomly from 3 to 7 math operation/letter strings, and participants had to complete 2 trials of each set size for a total possible score of 50. At recall for each set, letters from the corresponding set had to be recalled in order by selecting the relevant letters. See Unsworth, Heitz, Schrock, and Engle (2005) for more details.

Symspan. Participants solved symmetry judgements while remembering the location of a sequence of red squares within a matrix. Symmetry judgements consisted of an 8×8 matrix of squares in which some of the squares were filled black and the remaining squares remained white. Participants indicated whether the pattern created by the filled squares was symmetrical about the vertical axis. Once participants indicated whether they believed the pattern was symmetrical or non-symmetrical, participants were shown a 4×4 matrix with one of the cells filled red for 650 ms. Immediately following the presentation of the red square matrix, the next symmetry judgement trial began. Set sizes randomly ranged from 2 to 5, and there were 2 trials of each set size for a total possible score of 28. Participants were asked to recall the sequence of red-square locations based on the order in which they were presented across the corresponding trial. Participants indicated the appropriate location of each red-square by clicking on cells in an empty matrix. See Unsworth, Redick, Heitz, Broadway, and Engle (2009) for more details.

Rspan. While remembering the same unrelated letters as in the Ospan, participants provided judgements about a series of sentences. More specifically, participants read a sentence containing 10 to 15 words and determined whether or not the sentence made sense to them (e.g., "Every now and then I catch myself swimming blankly at the wall"). Nonsense sentences were created by modifying a single word from an otherwise ordinary sentence (e.g., changing "staring" to "swimming" in the case above). Upon indicating whether the sentence made sense or not, participants were then presented with a letter for 1 s. Set sizes randomly varied from 3 to 7 sentence/letter strings, and participants had to complete 2 trials of each set size for a total possible score of 50. At recall for each set, letters from the corresponding set had to be recalled in order by selecting the appropriate letters. See Unsworth et al. (2009) for more details.

Factor WM Score. All complex span tasks showed large intercorrelations: Ospan correlated with Rspan ($r = .62$, $p < .001$) and Symspan ($r = .40$, $p < .001$), and Rspan correlated with Symspan ($r = .39$, $p < .001$). All analyses involving WM used a WM factor score created for each participant by entering scores on the three complex span measures into a factor analysis using principal axis factoring. Factor loadings for the first unrotated factor were as follows: Ospan (0.76), Symspan (0.49), and Rspan (0.81). This variable allowed us to treat WM as a continuous variable in all analyses.

LTM tasks

Delayed Free Recall. Participants were administered a delayed free recall task consisting of 5-word lists containing 10 words each. Word lists were initially composed of randomized nouns selected from the Toronto word pool (Friendly, Franklin, Hoffman, & Rubin, 1982). All participants received the same lists of words and were instructed to recall as many words as possible from each list. Words were presented onscreen for 1 s, with each word preceded and followed by a 500 ms blank screen. Following presentation of the last word within each list, a 16 s distractor task began that required participants to verbally report a series of 8 three-digit numbers in descending order (adapted from Rohrer & Wixted, 1994). Each 3-digit string was presented onscreen for 2 s. After the distractor task, 3 question marks appeared in the center of the screen to prompt participants to recall as many words as possible within a 45 s window. Participants typed their responses in any order they wished and pressed "enter" after each word, thereby clearing the screen. A participant's score was proportion of items recalled correctly.

Picture Source Recognition. During the encoding phase, participants were presented with a picture (30 total pictures) in one of four different quadrants onscreen for 1 s. Participants were explicitly instructed to pay attention to both the picture (item) as well as the quadrant it was located in (source). At test, participants were presented with 30 old and 30 new pictures in the center of the screen. Participants were required to indicate if the picture was new or if it was old. If the picture was deemed old, they also had to specify what quadrant the picture was presented in via key press. Thus, on each test trial participants pressed one of five keys indicating new, old-top left, old-top right, old-bottom left, or old-bottom right. Participants had 5 s to press the appropriate key to enter their response. A participant's score was the proportion of correct responses.

Factor LTM Score. Given our interest in examining LTM ability common to various LTM tasks—not just paired associate learning—we used delayed free recall and source memory tasks to create a LTM factor composite. Consistent with prior work (Miller et al., 2019; Miller & Unsworth, 2020), scores on these two tasks were correlated ($r = .34$, $p < .001$) and entered into a factor analysis using principal axis factoring to create a LTM factor score for each participant. The factor loadings for the first unrotated factor were as follows: delayed free recall (0.59) and picture source recognition (0.59).

Paired associates cued-recall task

Participants were administered 3 lists of 30 word-pairs each. Similar to the delayed free recall task, word-pair lists were composed of randomized common nouns selected from the Toronto word pool (Friendly

Table 1
Descriptive statistics and reliability estimates for all measures in Experiment 1.

Measure	*N*	*M*	*SD*	Skew	Kurtosis	Reliability
Ospan	133	36.55	8.84	−.93	.86	.72
Rspan	135	34.74	8.69	−.60	−.02	.70
Symspan	134	18.96	4.91	−.23	−.22	.55
PAacc	137	.24	.20	1.11	.70	.84
DFRacc	134	.44	.17	.67	.61	.83
PicSource	128	.70	.20	−.88	.03	.95
PropOnTask	137	.36	.36	.49	−1.28	.84
PropTRI	137	.27	.25	.96	.31	.64
PropOffTask	137	.37	.36	.54	−1.22	.87
Motivation	137	3.39	1.52	−.17	−1.07	

Note. Due to program or experimenter error, some tasks are missing data. Ospan = operation span, Rspan = reading span, Symspan = symmetry span, PAacc = paired associates cued recall accuracy, DFRacc = delayed free recall accuracy, PicSource = picture source recognition accuracy, PropOnTask = proportion of on task thought, PropTRI = proportion of task-related interference, PropOffTask = proportion of off task thought (i.e., attentional lapses). All reliabilities were calculated using Cronbach's alpha.

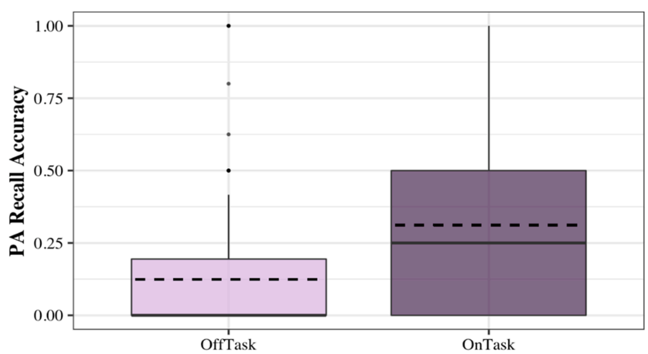

Fig. 1. A standard boxplot displaying recall accuracy for items immediately preceding off task (mind wandering, external distraction, and mind-blanking) and on task reports to thought-probes. Dotted lines reflect the mean for each thought type. Dots reflect outliers.

et al., 1982), and all words were between 3 and 6 letters in length. The task began with a "Ready?" signal onscreen, at which point participants pressed the spacebar to begin. Each list began with the same "Ready?" signal. Word pairs were preceded and followed by a blank screen onscreen for 500 ms, and word-pairs were presented vertically for 2 s each. All word pairs were associatively and semantically unrelated. Participants were told that the cue would always be the word on top and the target would be on the bottom. After the presentation of the last word pair, participants saw the cue word and ??? in place of the target word. Participants were instructed to type in the target word from the current list that matched cue. Consistent with prior work similarly administering long word-pair lists (e.g., Bailey, Dunlosky, & Kane, 2008), cues for the corresponding target words were presented in the same order at recall as they were presented during the encoding phase. Participants had 5 s to type in the corresponding word. A participant's score was the proportion of items recalled correctly.

Thought Probes. Probes pseudo-randomly appeared during the encoding phase of each word-pair list. Three probes appeared in list 1, four probes in list 2, and three probes in list 3. Each probe asked participants to report the current contents of their thoughts. Specifically, a screen appeared instructing participants to "Please characterize your current conscious experience." Response options were (1) I am totally focused on the current task, (2) I am thinking about my performance on the task, (3) I am distracted by sights/sounds/physical sensations, (4) I am daydreaming/my mind is wandering about things unrelated to the task, and (5) I am not very alert/my mind is blank. Participants responded by pressing the appropriate number on the keyboard. Response 1 was considered as on task, whereas response 2 was considered task-related interference (TRI)—instances in which thoughts are focused internally but are related to the appraisal of the current task. This option was included because prior work (Stawarczyk et al., 2011) indicates TRI is not a complete off task state, but nor is it considered a state where attention is entirely focused on the task being performed (i.e., TRI is stimulus-independent whereas on task processing is stimulus dependent). Finally, responses 3–5 were considered off task, aka attentional lapses (external distraction, mind-wandering, and mind-blanking).

Motivation. Upon completion of the PA task, participants were asked about their motivation during the task (Robison & Unsworth, 2018; Robison et al., 2020; Unsworth & McMillan, 2013). Specifically, participants were asked, "How motivated were you to perform well on the task?" Participants responded using a 6-point scale (1 = "Not at all motivated"; 6 = "Extremely motivated").

Results and Discussion

Table 1 reveals that all measures displayed adequate variability and were approximately normally distributed. Participants reported being on task 36% and off task 37% of the time (the remaining time was spent thinking about their performance). Logistic multi-level modeling (MLM) was then used to examine differences in subsequent recall as a function of attentional state (Off task, On task, and TRI)[2]. We opted to use MLM techniques here because a mean-based analytic technique, such as repeated measures ANOVA, only uses data from participants with complete data. Analyzing the data with a repeated measures ANOVA would have left us with a total sample size of 54, because some participants never reported being off task (mind-wandering, mind-blanking, or succumbing to external distraction), some never reported being on task, and some never reported an instance of TRI. Using MLM, we were able to leverage observations from participants that would have been excluded. With respect to the model itself, to-be-remembered items were nested within probe number (10 total thought-probes each corresponding to a specific to-be-remembered item) and subjects. In other words, probe number and subject were specified as random effects (i.e., intercepts were allowed to vary across subjects and probes), and 10 outcomes per subject were included in the analysis. The outcome reflects recall (recalled vs forgotten) of the word-pair immediately preceding each of the 10 thought-probes, meaning only 1/9th of each subjects' recall performance was analyzed. Our fixed effect was attentional state. The reference group was off task thought, but results remained unchanged when specifying on task thought as the reference group. Results revealed on task thought ($\gamma = 1.48$, $SE = .24$, z-value = 6.23, $p < .001$) was associated with better expected log odds of subsequent recall relative to off task thought. In other words, the odds of a participant correctly recalling an item were 4.39 ($e^{1.48}$) times greater when they reported being on task compared to when they were off task. Fig. 1, a standard boxplot, more clearly demonstrates this effect: subsequent memory was more accurate when individuals reported being on task than off task. An additional finding gathered from Fig. 1 is that—despite the overall proportion of off task thought (37%) being nearly identical to the overall proportion of on task thought (36%)—50% of the subjects experiencing off task thoughts provided no correct retrievals for the studied items preceding these off task reports. Taken altogether, these results are consistent with prior research (deBettencourt, Norman, & Turk-Browne, 2018; Garlitch & Wahlheim, 2020; Maillet & Rajah, 2013,

[2] We had no specific hypotheses about how TRI might relate to recall performance. We were primarily interested in examining differences between on task thought and off task thought. Nonetheless, results revealed TRI ($\gamma = .59$, $SE = .24$, z-value = 2.49, $p = .01$) was associated with better expected log odds of subsequent recall relative to off task thought. The results reported within the main text remained unchanged when excluding reports of TRI.

Table 2
Correlations among all measures in Experiment 1.

Measure	1	2	3	4	5
1. PAacc	–				
2. WM	.18*	–			
3. LTM	.61***	.34***	–		
4. PropOffTask	−.35***	−.07	−.29**	–	
5. Motivation	.52***	−.001	.37***	−.57***	–

Note. * $p<.05$, ** $p<.01$, *** $p<.001$; PAacc = paired associates cued recall accuracy, WM = working memory factor composite, LTM = long-term memory ability factor composite, PropOffTask = proportion of off task thoughts (i.e., attentional lapses).

Table 3
Simultaneous regression predicting PA cued recall accuracy in Experiment 1.

Variable	N	β	t	sr^2	R^2	F
LTM	127	.47	6.03***	.16		
WM	131	.01	.19	.00		
PropOffTask	137	−.03	−.34	.001		
Motivation	137	.33	3.85***	.07	.47	26.11

Note. * $p<.05$, ** $p<.01$, *** $p<.001$; Participants with missing data were excluded from this analysis via pairwise deletion. WM = working memory factor composite, LTM = long-term memory ability factor composite, PropOffTask = proportion of off task thoughts (i.e., attentional lapses).

2014; Metcalfe & Xu, 2016; Seibert & Ellis, 1991; Smallwood, Baracaia, Lowe, & Obonsawin, 2003; Thomson, Smilek, & Besner, 2014; Xu & Metcalfe, 2016) suggesting that when participants experience a lapse of attention during learning, they are less likely to remember that information at test.

Next, we turn to our individual differences analyses. All correlations and regressions used the full, usable sample, and each participant contributed one observation for each variable in the analysis. In cases where there was missing data, pair-wise deletion was used. Consistent with prior work (Garlitch & Wahlheim, 2020; Metcalfe & Xu, 2016; Seibert & Ellis, 1991; Thomson, Smilek, & Besner, 2014; Xu & Metcalfe, 2016), Table 2 reveals a negative correlation between off task thoughts and overall PA accuracy ($r = -.35$, $p < .001$), suggesting that participants who experienced more lapses of attention (less consistency) were less able to sufficiently learn the PA task. While proportions of off task thought were unrelated to WM, those who reported more off task thoughts at encoding tended to have worse LTM ability ($r = -.29$, $p < .01$). These individuals who struggled to consistently keep attention on task also tended to report lower levels of motivation to do well on the PA task ($r = -.57$, $p < .001$). Increased motivation was, in turn, strongly associated with better recall accuracy ($r = .52$, $p < .001$).

Given the strong association between consistency and motivation, we next sought to directly test the notion that motivation (in conjunction with WM and LTM abilities) accounts for the observed relationship between consistency and learning. That is, we examined whether variation in the consistency of attention is still predictive of learning ability when taking these other factors into account. As seen in Table 3, WM, LTM ability, proportions of off task thought, and motivation together accounted for 47% of the variance in PA recall accuracy. Critically, with motivation as a predictor, off task thoughts did not account for unique variance in learning ability (due to substantial shared variance between the two constructs). Indeed, when dropping motivation as a predictor from the simultaneous regression model, the consistency of attention, as indexed by proportions of off task thought, now accounted for unique variance in recall performance ($\beta = -.19$, $sr^2 = .03$, $p < .05$). So, while individual differences in the consistency of attention seem to be important in accounting for variation in learning ability, these results suggest that the relationship between these two factors is entirely driven by motivation, insofar that attention is less consistently allocated to on task processing among those who are less motivated to perform well, resulting in overall worse learning (see Seli et al., 2019). [3]

Experiment 2

Experiment 1 revealed that subsequent memory was worse on trials when participants reported having an attentional lapse relative to when participants reported being on task. A large (Funder & Ozer, 2019; Gignac & Szodorai, 2016) negative correlation also emerged between lapses of attention and PA recall accuracy ($r = -.35$), suggesting that the consistency of attention is an important predictor of associative learning abilities. That said, Experiment 1 further demonstrated that proportions of off task thought did not account for any unique variance in learning ability when motivation was included as a predictor in the simultaneous regression model, suggesting the relationship between the consistency of attention and learning ability was entirely determined by differences in motivation. Note, however, that the *zero-order* correlation between motivation and proportions of off task thought in Experiment 1 ($r = -.57$) was of a similar magnitude to correlations typically observed at the *latent* level (latent $r = -.52$, see Appendix A in Robison et al., 2020). It therefore seems possible that our use of a single self-report item following task completion may have overestimated the effect of motivation on both consistency and learning. That is, when motivation is only measured after completing a single task, it is unclear how performance and/or attentional lapse rates during the task may reactively influence these reports. Motivation and other state-based variables should ideally be measured both before and after the criterion task to reduce potential reactivity effects (via calculation of mean motivation scores or examination of intercepts and slopes). To address this concern, the first aim of Experiment 2 was to replicate and extend Experiment 1′s motivation-related effects on consistency and learning by including a second motivation self-report item that appeared upon task onset (i.e., immediately before beginning the real trials).

The second aim of Experiment 2 was to better elucidate how intensity is related to overall learning abilities. As described in the Introduction, motivation could similarly account for differences in intensity. Individuals who are more motivated to perform well may be more inclined to put forth extra attentional effort during encoding in an attempt to maximize their performance. Based on this account, motivation and TEPRs at encoding should positively correlate, and motivation should also explain the positive correlation between TEPRs and PA recall accuracy. Furthermore, variation in intensity is also likely influenced by an individual's overall attentional resource capacity (total amount of potential attention to allocate). Some individuals will simply have more available attentional resources than others. If one were to assume that the capacity of WM is synonymous with the capacity of attentional resources, then we would expect WM to positively correlate with TEPRs at encoding and to also account for the relationship between TEPRs and PA recall accuracy. Yet another possibility is that variation in intensity may also be explained by differences in consistency. Theoretically, less attention is allocated to learning during a lapse compared to when participants are completely focused on the task, because attention is at

[3] A simultaneous regression without the LTM composite (i.e., WM, off task thoughts, and motivation are the only predictors), yielded the same pattern of results except WM became a significant unique predictor of recall accuracy. The beta estimate for off task thought became numerically stronger (β =-.06) but still did not account for significant unique variance in recall accuracy (p = .479). Hence it appears that motivation is the primary factor responsible for the finding that consistency does not uniquely predict associative learning in Experiment 1. When repeating this analysis for Experiment 2, WM became a marginally significant (p = .052) unique predictor of recall accuracy. The beta estimates for off task thought (β = -.31) and TEPRs (β = .28) also became numerically stronger. The beta estimate for motivation increased (β = .10) but remained non-significant (t = 1.22, p = .22).

Table 4
Descriptive statistics and reliability estimates for all measures in Experiment 2.

Measure	*N*	*M*	*SD*	Skew	Kurtosis	Reliability
Ospan	164	37.23	8.47	−.96	1.15	.68
Rspan	166	37.39	8.40	−.84	.41	.73
Symspan	166	18.71	5.23	−.62	−.24	.63
PAacc	165	.30	.22	.81	−.33	.88
DFRacc	166	.51	.13	.11	.82	.53
PicSource	166	.76	.17	−1.10	.71	.93
PropOnTask	165	.30	.32	.89	−.42	.85
PropTRI	165	.32	.24	.92	.77	.68
PropOffTask	165	.39	.29	.32	−.90	.76
MeanMotivation	165	4.35	1.07	−.15	−.75	.79
TEPR	148	.06	.06	.76	1.90	.96

Note. Ospan = operation span, Rspan = reading span, Symspan = symmetry span, PAacc = paired associates cued recall accuracy, DFRacc = delayed free recall accuracy, PicSource = picture source recognition accuracy, PropOnTask = proportion of on task thought, PropTRI = proportion of task-related interference, PropOffTask = proportion of off task thought (i.e., attentional lapses), MeanMotivation = mean motivation across pre and post measures, TEPR = mean task evoked pupillary response at the last encoding bin (the final 200 ms). The reliability for MeanMotivation was calculated using the split-half method on pre and post motivation scores. All other reliabilities were calculated using Cronbach's alpha. Two participants were missing Ospan data due to computer program malfunction. One other participant did not complete the PA task because they did not have time to complete it. Finally, 18 participants were missing TEPR data due to blinks, off-screen fixations, or eye-tracker malfunction.

least partially split between whatever is causing the lapse and learning the task at hand (i.e., perception becomes decoupled from the external stimulus). Accordingly, two predictions follow: (1) lapses within individuals should be associated with a momentary reduction in attentional intensity (smaller TEPRs), and (2) at the individual differences level, the ability to more consistently maintain attention on task (fewer off task thoughts) should be associated with increased intensity of attention to items (larger TEPRs) at encoding. In sum, we sought to examine whether TEPRs are associated with motivation, WM, and proportions of off task thought and also determine whether TEPRs uniquely predict associative learning when controlling for these factors.

The third and final goal of Experiment 2 was to more specifically address the question of whether intensity and consistency reflect different, unique abilities, or whether they reflect the same, general ability to control aspects of one's attention at learning. So, an important point of analysis was to first determine whether consistency and intensity both contribute unique variance in predicting learning ability when taking each other into account as well as other important variables, including motivation, WM, and general LTM abilities. If intensity and consistency both explain unique variance in learning ability (i.e., PA recall accuracy), we sought to further determine how intensity and consistency are related yet distinct. To address these questions, participants completed the same procedure as Experiment 1, except we added a second motivation self-report measure and pupillary responses were simultaneously recorded throughout the duration of the PA cued recall task to provide an index of the intensity of attention.

Method

A total of 171 participants (66.5% female) were recruited from the human subjects pool at the University of Oregon and were compensated with course research credit for participating. One individual was excluded from all analyses because they were not proficient in English, and scatterplots revealed them to be an outlier on multiple tasks. One other participant was excluded for being over 35 years of age, another for being an outlier on the PicSource (i.e., recall accuracy was below 10%), and two more for not being able to recall a single word-pair across three lists on the PA task (final $N = 166$). All remaining participants were between the ages of 18 and 30 ($M = 19.40$, $SD = 1.82$). With the exception of the addition of eye-tracking, the procedure was identical to Experiment 1. After obtaining informed consent and demographic information, all participants completed the Ospan, Symspan, and Rspan as measures of WM. Next, participants completed DFR and PicSource as our LTM battery. Upon completion of LTM tasks, participants were then moved to a dimly lit room where they completed the PA cued recall task while pupil diameter was simultaneously recorded binocularly at 120 Hz using a Tobii T120 eye-tracker.

Prior to beginning the PA task, participants were seated 60 cm from the monitor, and a 9-point standard calibration procedure began. To calibrate the eye tracker, participants were asked to fixate on a series of 9 grey dots presented on a white background. The Tobii Eye Tracker measures aspects of the participant's eyes and uses them together with an internal, anatomical 3D eye model to calculate the mapping between the identified gaze position on the display and the eye tracker's estimate of that position. Recalibration occurred whenever the criterion defined by the proprietary software was not met. All participants were successfully calibrated within the first few attempts. In addition, a headrest, mounted at the front of the table holding the eye-tracker, was used to reduce any potential influence of uncontrolled head movements. We should also note that participants completed the tasks reported herein as part of a larger experimental test battery lasting approximately 1.5 h. Since the other tasks administered during the experimental session do not relate to the current study, they are not reported.

WM tasks

Same as Experiment 1. Ospan correlated with Rspan ($r = .61$, $p < .001$) and Symspan ($r = .31$, $p < .001$). Rspan also correlated with Symspan ($r = .27$, $p < .001$).

LTM tasks

Same as Experiment 1. Mean accuracy on DFR and PicSource were correlated ($r = .35$, $p < .001$). The only difference was that the DFR task in Experiment 2 had to be shortened to 3 lists of 10 words (rather than 5 lists of 10 words) for the sake of time. As demonstrated in Table 4, shortening the task yielded lower but adequate reliability (Cronbach's $\alpha = .53$).

Paired associates cued-recall task

After calibration of the eye-tracker, participants were administered 3 lists of 30 word-pairs each. Word-pair lists were identical to those used in Experiment 1—words were composed of randomized common nouns selected from the Toronto word pool (Friendly et al., 1982). All words were between 3 and 6 letters in length. Words were not allowed to repeat across tasks, and words (as well as the mask preceding/following each word) were presented in black text in Courier New font (font size = 40) on a light grey background. Properties such as ambient light, screen brightness, contrast, etc. were held constant across all participants. The task began with a "Ready?" signal onscreen, at which point participants pressed the spacebar to begin. Each list began with the same "Ready?" signal. Word pairs were presented horizontally for 3 s each, and every word pair was preceded by a mask of five plus signs replacing each word that had previously been onscreen (2 s duration; see Fig. 2 for schematic of task). The encoding period as well as the masking period between presentation of word-pairs was increased in duration from Experiment 1, because pupillary responses are relatively slow, and this fixation period was used to baseline correct pupillary responses on a trial-by-trial basis (i.e., each word pair had its own baseline). All other parameters are identical to Experiment 1.

Thought Probes. See Experiment 1.

Motivation. The only difference between experiments was that participants were now asked about their level of motivation twice during the task. The first self-report item appeared after completing the practice

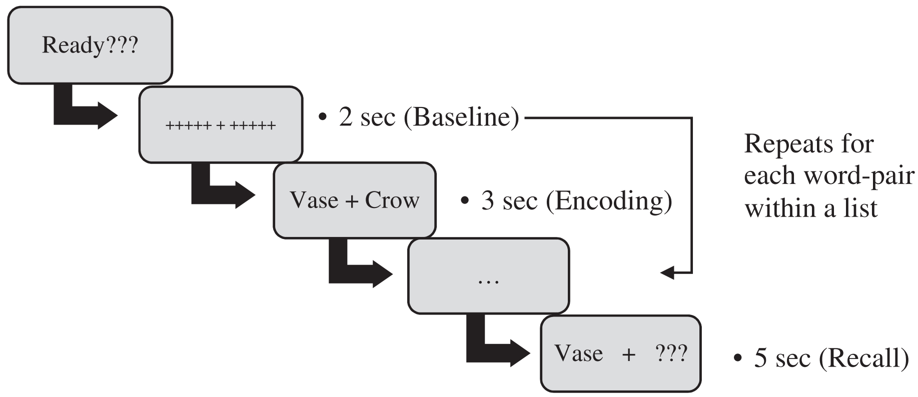

Fig. 2. Schematic of the paired associates (PA) cued-recall task.

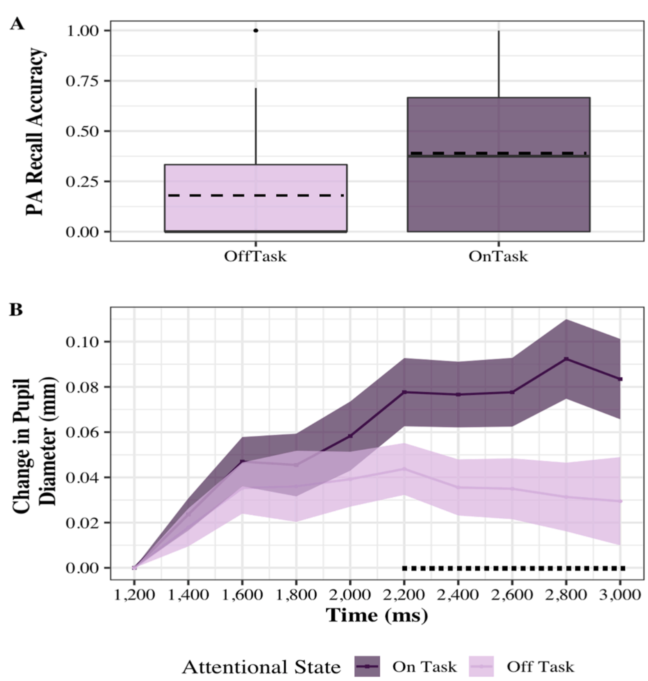

Fig. 3. (a) A standard boxplot displaying recall accuracy for items immediately preceding off task (mind wandering, external distraction, and mind-blanking) and on task reports to thought-probes. Horizontal dashed lines reflect the mean for each thought type. Dots reflect outliers. (b) Change in pupil diameter across the encoding period for each word-pair as a function of attentional state (on task vs off task). Shaded areas reflect one standard error of the mean. Horizontal dotted line illustrates time-points in which pairwise comparisons were significantly different when adjusting for multiple comparisons via the Bonferroni correction (p's $< .0056$).

trials, directly before beginning the real trials of the PA task. This item specifically asked participants to indicate how motivated they were to perform the upcoming real trials. The second self-report item appeared immediately upon completion of the PA task. The two measures of motivation were highly correlated ($r = .67$, $p < .001$), so we used the mean of these reports for our Experiment 2 analyses. See Experiment 1 for more details.

Task Evoked Pupillary Responses (TEPRs). As previously mentioned, pupil diameter was assessed continuously throughout the PA task. Data from each participant's left eye was used for analyses, and missing data points associated with eye tracker malfunction, blinks, or off-screen fixations were excluded from averaging (i.e., we did not interpolate missing pupil data). TEPRs were baseline corrected on a pair-by-pair basis by subtracting mean baseline diameter prior to word onset from the average pupil diameter during the 3 s encoding phase for each word. In addition, the pupil data for the 3 s encoding phase was broken down into a series of 200 ms timeframes, resulting in 15 total baseline corrected bins. However, consistent with prior work (Miller & Unsworth, 2020), TEPRs during the first 1200 ms were strongly confounded by changes in luminance. Following the appearance of the to-be-remembered word-pair onscreen, individuals tend to first briefly fixate on the Cue word (the word on the left side of the screen) and quickly alter their gaze to look at the Target word (the word on the right side of the screen). The appearance of the stimuli paired with more frequent gaze alternations seemingly results in large changes in luminance. Hence TEPRs during the first 1200 ms generally reflect changes in the pupillary light reflex (e.g., Binda, Pereverzeva, & Murray, 2013), not the intensity of attention. To better illustrate pupil dynamics related to the intensity of attention, the bin representing the end of the pupillary light reflex is used for baseline correction (see Wang, Brien, & Munoz, 2015 for similar method). In other words, like Miller and Unsworth (2020), we corrected TEPRs by using 1200 ms as our reference point, leaving a total of 9 time-bins for our TEPR-related analyses. We subtracted mean dilation at 1200 ms from each ensuing bin (see Appendix A for TEPR analyses across the entire 3 s encoding period). The dependent variable used for all correlation and regression analyses was the mean pupillary response at the last encoding bin (the final 200 ms) aggregated across all trials (including trials that were followed by thought probes).

Results and Discussion

All measures displayed adequate variability and were approximately normally distributed (i.e., skewness < 2; kurtosis < 4). Table 4 reveals that participants reported being on task 30% and off task 39% of the time; the remaining time was spent thinking about their performance. Replicating Experiment 1, the logistic MLM—with the same parameters and model structure specified in Experiment 1—revealed on task thought ($\gamma = 1.38$, $SE = .20$, z-value $= 6.92$, $p < .001$) was associated with better expected log odds of subsequent recall relative to off task thought[4]. Put another way, the odds of a participant correctly recalling an item were 3.96 ($e^{1.38}$) times greater when they reported being on task compared to when they were off task. Thus, as demonstrated in Fig. 3a, participants were less likely to remember an item at test if they experienced a lapse of attention during learning of said item. Fig. 3a further demonstrates that despite the overall proportion of off task thought again being similar to the overall proportion of on task thought, 50% of the subjects experiencing off task thoughts provided no correct retrievals for the studied items preceding these off task reports. Memory failures are evidently associated with the occurrence of attentional lapses at

[4] TRI was again associated with better expected log odds of subsequent recall relative to off task thought ($\gamma = 0.89$, $SE = .18$, z-value $= 4.87$, $p < .001$).

Table 5
Correlations among all measures in Experiment 2.

Measure	1	2	3	4	5	6
1. PAacc	–					
2. WM	.12	–				
3. LTM	.48***	.25**	–			
4. PropOffTask	−.44***	−.01	−.16*	–		
5. MeanMotivation	.32***	−.02	.13	−.53***	–	
6. TEPR	.37***	−.10	.23**	−.27**	.18*	–

Note. * $p<.05$, ** $p<.01$, *** $p<.001$; PAacc = paired associates cued recall accuracy, WM = working memory factor composite, LTM = long-term memory ability factor composite, PropOffTask = proportion of off task thoughts (i.e., attentional lapses), TEPR = mean task evoked pupillary response at the last encoding bin (the final 200 ms).

Table 6
Simultaneous regression predicting PA cued recall accuracy in Experiment 2.

Variable	*N*	*β*	*t*	sr^2	R^2	*F*
LTM	164	.37	5.20***	.12		
WM	162	.04	.58	.001		
PropOffTask	165	−.29	−3.63***	.06		
MeanMotivation	165	.08	1.06	.005		
TEPR	148	.19	2.75**	.03	.40	18.82***

Note. * $p<.05$, ** $p<.01$, *** $p<.001$; Participants with missing data were excluded from this analysis via pairwise deletion. WM = working memory factor composite, LTM = long-term memory ability factor composite, PropOffTask = proportion of off task thoughts (i.e., attentional lapses), TEPR = mean task evoked pupillary response at the last encoding bin (the final 200 ms).

learning.

Next, we sought to examine whether attentional lapses were associated with temporarily reduced intensity within individuals. That is, when participants report having a lapse of attention, do they have smaller TEPRs throughout the encoding period than they do when they report being on task? Multi-level modeling (MLM) was also used to examine differences in TEPRs at encoding as a function of attentional state (on vs off task; see Table 4 for proportions of thought types). So, similar to the recall analysis, we analyzed TEPRs for word-pairs appearing immediately before a thought-probe. Using a repeated measures ANOVA produced similar results but would have left us with a total sample size of 61. Using MLM, we were able to leverage observations from participants that would have been excluded. With respect to the model itself, TEPRs were nested within subjects. Our fixed effects included the linear effect of bin, the quadratic effect of bin, attentional state (the reference group was on task thought, but results remained unchanged when specifying off task thought as the reference group), and the interaction between the linear effect of bin and attentional state. Intercepts were allowed to vary across subjects. Also note that since TEPRs were baselined by the value at 1200 ms, all TEPR values corresponding to this 1200 ms bin were zero; hence we eliminated these values from the MLM analysis. Thus, for bin-related analyses on TEPRs, the beginning trial refers to dilations observed at the next ensuing bin, which was 1400 ms.

Results revealed that TEPRs at the start of the encoding period (i.e., at 1400 ms) did not significantly vary as a function of attentional state ($p > .64$). Critically, though, a significant interaction between bin and attentional state emerged, suggesting that the degree to which TEPRs increased across the encoding period depended on whether one was on task or experiencing an attentional lapse. As shown in Fig. 3b, TEPRs at encoding increased most when participants reported being on task [$b = .02$, $SE = .004$, $t(1{,}728.70) = 3.76$, $p < .001$]. While TEPRs corresponding to reports of off task thoughts mimicked on task thoughts early in the encoding period (such that TEPRs increased initially), this increase was smaller than what was observed for on task thoughts [$b = -.008$, $SE = .002$, $t(1{,}728.76) = -3.49$, $p < .001$]. Note that we also tested a model including the interaction between the quadratic effect of bin and attentional state, but model comparisons revealed this model did not significantly improve model fit [$\chi^2(1) = .05$, $p > .81$]; the quadratic trend did not significantly vary across attentional states ($p > .81$). Hence TEPRs for both on and off task thoughts tended to reach asymptotic levels at the same rate ($b = -.001$, $SE = .000$, $t(1{,}728.78) = -2.15$, $p < .05$).[5]

The results above suggest that intensity and consistency are linked within participants, such that less attention is allocated to learning (intensity is reduced) during a lapse relative to when participants are fully focused on task[6]. Examining individual differences revealed a similar trend (see Table 5). TEPRs negatively correlated with proportions of off task thought ($r = -.27$, $p < .01$), suggesting that those who allocated more attentional intensity to items were also better able to consistently keep attention on task. Thus, the heightened state of task readiness/task engagement associated with increased intensity (Kahneman, 1973) seems to make it more difficult for potent task-irrelevant concerns to break into the focus of attention. Compatible with Experiment 1, lower proportions of off task thought (more consistency) during encoding were unrelated to WM but were weakly associated with better LTM abilities ($r = -.16$, $p < .05$). Fewer off task thoughts were also strongly associated with higher *mean* levels of motivation ($r = -.53$, $p < .001$) and superior PA recall accuracy ($r = -.44$, $p < .001$). Turning to correlations more specific to intensity, Table 5 reveals no significant correlation between TEPRs and WM, but larger TEPRs were associated with enhanced LTM ability ($r = .23$, $p < .01$), higher motivation ($r = .18$, $p < .05$), and better PA recall accuracy ($r = .37$, $p < .001$). Thus, consistent with Miller and Unsworth (2020), the best learners devoted more attention (greater intensity) to items during learning. These high intensity individuals also tended to be high in consistency, have higher levels of motivation, and have better LTM abilities in general.

Collectively, the results thus far suggest that the intensity and consistency of attention are related to each other in addition to other important individual differences variables such as motivation and (to a lesser extent) general LTM abilities. To better assess whether these two aspects of attention merely reflect the same general ability to control attention during learning, we next examined whether the intensity and consistency of attention explain unique variance in learning above and beyond what is accounted for by each other and these other variables. Proportion of off task thoughts, TEPRs at encoding, WM factor composites, LTM factor composites, and mean motivation scores were all added to a simultaneous linear regression model predicting mean PA recall accuracy—while WM did not significantly correlate with PA recall accuracy, we added WM to the simultaneous regression model to maintain consistency with Experiment 1. As demonstrated in Table 6, the predictors accounted for 40% of the variance in recall accuracy.

[5] Allowing slopes to vary for attentional state resulted in a significantly better fitting model than the one reported, χ^2 (2) = 464.71, $p < .001$, but results remained the same: TEPRs at the beginning of the encoding period did not differ as a function of attentional state ($p > .93$), and a large increase in TEPRs was observed across the encoding period when participants reported being on task [$b = .02$, $SE = .003$, $t(1{,}659.39) = 4.59$, $p < .001$]. Critically, the increase in TEPRs for off task reports was significantly smaller than what was observed for on task reports [$b = -.008$, $SE = .002$, $t(1{,}659.41.41) = -4.32$, $p < .001$].

[6] We also examined subsequent memory effects, irrespective of on vs off task thought, to see whether TEPRs (assessed at the final encoding bin) would predict subsequent recall within individuals. Using logistic MLM, TEPRs were entered as a fixed effect, Subject and Trial were random effects, and intercepts were allowed to vary. Results revealed TEPRs were significant predictors of subsequent recall ($\gamma = 0.45$, $SE = 0.11$, z-value $= 4.02$, $p < .001$). For every one-mm increase in pupil dilation at the final encoding bin, the odds of an individual correctly recalling said item were 1.57 ($e^{0.45}$) times greater. Put more simply, recalled items ($M = .086$, $SE = .004$) were associated with larger TEPRs than forgotten items ($M = .047$, $SE = .003$).

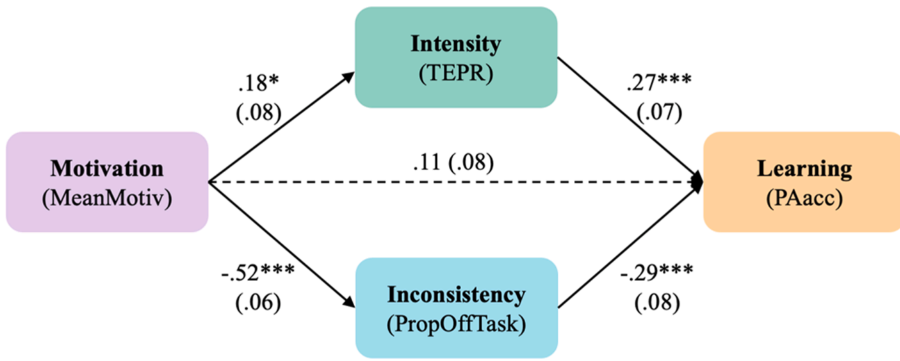

Fig. 4. * $p<.05$, ** $p<.01$, *** $p<.001$; Path model in which motivation (MeanMotiv = mean motivation scores) predicts intensity (TEPR = mean task evoked pupillary response during the final encoding bin) and inconsistency (PropOffTask = proportions of off task thought); and motivation, intensity, and inconsistency all predict learning (PAacc = paired-associate cued recall accuracy). Single-headed arrows connecting manifest variables (rectangles) to each other represent standardized path coefficients, indicating the unique contribution of the manifest variable. Numbers in parentheses reflect the standard error around each estimate. Solid lines are significant at the $p < .05$ level, and dotted lines are not significant at the $p < .05$ level. Intensity and inconsistency were allowed to covary in the model ($r = -.21$, $SE = .08$, $p < .01$), but this covariance was not added to the figure to maintain simplicity.

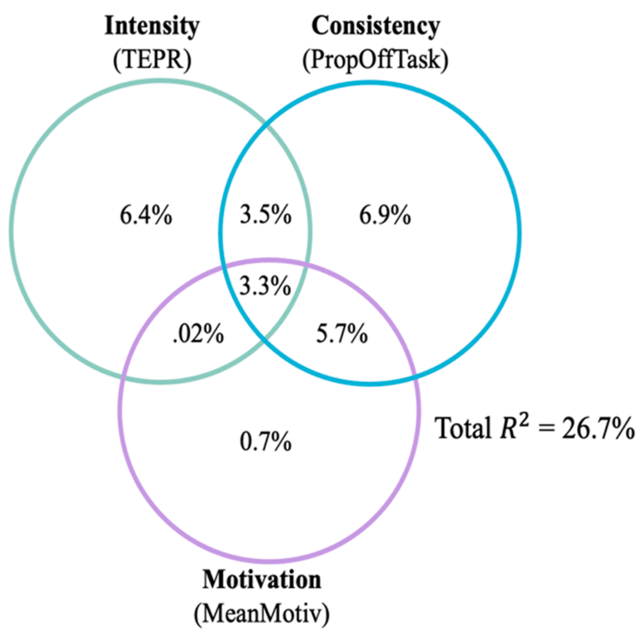

Fig. 5. Venn diagrams representing the shared and unique variance between the intensity of attention (mean TEPR during the final encoding bin), the consistency of attention (proportion of off task thought), and motivation (mean motivation) in predicting associative learning ability (PA cued recall accuracy).

Unsurprisingly, higher scores on the composite LTM variable were associated with better recall accuracy on the PA task. Critically, though, even when controlling for the influence of these more general LTM abilities in addition to mean motivation and WM, the consistency of attention (as indexed by proportions of off task thought) now accounted for unique variance in recall accuracy. In a related vein, the intensity of attention (as indexed by TEPRs) also accounted for unique variance in recall accuracy. One concern, of course, is that perhaps analyzing the data similar to Experiment 1—with post-motivation scores as opposed to mean motivation scores—could yield a different pattern of results for consistency. Notably, the results remain the same when substituting post-motivation scores for mean motivation scores (see Appendix B for results with pre- and post-motivation scores in isolation). This is a result at odds with Experiment 1, which will be discussed more in the General Discussion. Nonetheless, while the intensity and consistency of attention are (1) linked both within and between individuals and (2) related to motivation and general LTM abilities, each are seemingly distinct aspects of attention that are uniquely predictive of associative learning ability.

To more formally test the notion that intensity and consistency are distinct aspects of attention that are differentially influenced by various factors, we next specified a path model where mean motivation scores (motivation) predicted pupillary responses (intensity) and proportions of off task thought (inconsistency). Mean motivation, pupillary responses, and proportions of off task thought all predicted PA recall accuracy (learning ability). Pupillary responses and proportions of off task thought were also allowed to covary. As noted previously, individuals who allocated more attentional effort to learning appeared to be more motivated (*r* between TEPR and mean motivation = .18). That is, people who were more motivated to perform well were seemingly more willing to mobilize extra effort to enhance performance. If *learning*-related differences in intensity are partly explained by motivation, this account predicts a positive direct effect of the intensity of attention on learning, but this relation should be partially due to variation in motivation (an indirect effect). Similarly, individuals who were more motivated appeared to more consistently allocate attention to task-relevant information across trials (*r* between off task thoughts and mean motivation = −.53). If *learning*-related differences in consistency are partly explained by motivation, this account predicts a negative direct effect of inconsistency on learning (greater proportions of off task thought associated with worse learning), which should also be partially due to variation in motivation (an indirect effect). In sum, if intensity and consistency are distinct aspects of attention differentially impacted by various factors, both should have direct effects on learning even when accounting for the influence of motivation (which should have differential influences on intensity and consistency).

Shown in Fig. 4 is the resulting model. Mean motivation positively predicted TEPRs at encoding and negatively predicted the occurrence of off task thought during learning. Both increased TEPRs (greater intensity) and fewer off task thoughts (more consistency) had significant direct effects on learning. The direct effect of motivation on learning was not significant when taking the intensity and consistency into account, although there was a strong indirect effect of motivation on learning through consistency ($\beta = .15$, $p = .002$). Interestingly, the indirect effect of motivation on learning through intensity did not reach conventional levels of significance ($\beta = .05$, $p = .06$). So, while higher motivation is associated with increased intensity, the mobilization of attentional effort that is solely associated with increased motivation does not appear to impact learning performance. Motivation instead appears to primarily exert its influence on learning by determining how attention is allocated to various on versus off task activities across trials (consistency). These results support the notion that the intensity and consistency of attention have distinct, direct effects on learning, and each are differentially impacted by motivation.

Finally, in an effort to further disentangle how intensity and consistency are similar yet distinct, we utilized variance partitioning methods (e.g., Chuah & Maybery, 1999; Cowan et al., 2005) to distribute the overall R^2 of PA recall accuracy into portions shared and unique to TEPRs, proportions of off task thoughts, and mean motivation. A series of regression analyses were used to obtain R^2 values from each of the predictors in order to partition the variance. For each variable entering the regression, the zero-order correlations from Table 5 were used, and participants with missing data were excluded from the analysis via pairwise deletion. Fig. 5 illustrates that, altogether, TEPRs, off task thoughts, and motivation explained 26.7% of the variance in associative learning ability. Of this variance, only 3.3% was shared by all three constructs, whereas the remaining 23.4% was attributed to both unique and shared variance across the three constructs. The 3.3% of variance shared by all three constructs likely represents the lowest (or highest) performers—low learning individuals who devoted less attentional effort to learning who were likewise more susceptible to off task thoughts/lapses of attention (low intensity + low consistency) because they were less motivated to perform well. Interestingly, the intensity and consistency of attention also shared 3.5% of variance independent of motivation, meaning some low-learning individuals devoted less attention to items at encoding and experienced more attentional lapses (low intensity + low consistency) despite being highly motivated.

Consistent with the path model above, we see that when controlling for variance shared with off task thought, TEPRs and motivation essentially shared no variance (0.2%) when predicting learning ability. On the other hand, when partialling out variance shared with TEPRs, off task thoughts continued to share 5.7% of variance with motivation. Hence motivation appears to play a large role (independent of intensity) in determining how consistently attention is allocated to on vs off task processing. Critically, though, TEPRs and proportions of off task thought each contributed considerable unique variance in learning ability (6.4% and 6.9%, respectively), suggesting that both aspects of attention explain substantial differences in learning independent of each other and motivation. Namely, irrespective of one's motivation to perform well, some individuals demonstrated high levels of performance when they were on task because they were able to allocate a large amount of attentional effort to the study material, but these individuals had many trials with lower performance due to frequent lapses of attention (high intensity + low consistency). Conversely, some individuals allocated low levels of attentional effort to the current task but their allocation of attention to on task processing did not change much from trial-to-trial (low intensity + high consistency). From these results, it is clear that various learning profiles exist that are specific to the intensity and consistency of attention. While motivation explains some of this variation, future research needs to better identify the variables responsible for the remaining unique variance explained by intensity and consistency.

In sum, Experiment 2 was consistent with Experiment 1 insofar that the best learners were both more motivated and better able to consistently maintain attention on task. Moreover, motivation and consistency (as indexed by proportions of off task thought) were once again strongly positively correlated despite the use of mean motivation scores instead of post motivation scores. Unique to Experiment 2 and consistent with Miller and Unsworth (2020; see also Miller et al., 2019), the best learners also displayed larger TEPRs during encoding, suggesting that high learning individuals devoted more attentional effort (intensity) to items during encoding. An additional, novel finding was that TEPRs negatively correlated with proportions of off task thought and positively correlated with motivation; hence high intensity individuals tended to be high in consistency and high in motivation. Critically, though, while the intensity and consistency of attention were clearly connected—both within and between individuals—regression analyses showed that TEPRs (intensity) and proportions of off task thought (inconsistency) both accounted for unique variance in learning when accounting for each other and other meaningful predictors (general episodic memory abilities, motivation, and WM). Finally, analyses further revealed that while the intensity and consistency of attention were both related to motivation and (to a lesser extent) general LTM abilities, motivation appeared to be a factor more so implicated in the consistency of attention. Specifically, when controlling for variance shared with off task thoughts, TEPRs were no longer related to motivation. Conversely, even when controlling for shared variance with TEPRs, the ability to consistently prevent attentional capture from task-irrelevant information during learning (more consistency) was largely—but not entirely—the result of increased motivation. Taken altogether, Experiment 2 supports the notion that the intensity and consistency of attention are likely distinct, multifaceted constructs influenced by a variety of slightly different factors.

General discussion

The present study had four over-arching aims. Specifically, we sought to better understand why (1) the ability to consistently maintain attention on task (consistency) during learning and (2) the ability to allocate large amounts of attention (intensity) to the to-be-remembered material predict learning ability on a PA cued recall task. Considering that the existing literature has yet to simultaneously examine both consistency and intensity, we (3) also wanted to determine whether these two aspects of attention reflect different, unique abilities or whether they reflect the same, general ability to control aspects of one's attention at learning. Finally, (4) if intensity and consistency explained unique sources of variance in learning, we aimed to better understand how these aspects of attention are related yet distinct.

In regard to the first aim, both experiments revealed that worse learners tend to be individuals who less consistently allocate attention to on task processes during learning. Thus, several recent studies (Garlitch & Wahlheim, 2020; Metcalfe & Xu, 2016; Thomson, Smilek, & Besner, 2014; Xu & Metcalfe, 2016) have demonstrated a reliable negative correlation between lapses of attention/off task thoughts and learning at the individual differences level. We extended this work by first showing that those who lapsed the most (i.e., those with less consistency) during encoding also tended to have lower motivation and worse LTM abilities in general. The relationship between motivation and proportions of off task thought at encoding was especially strong (*r*s in both experiments $> .50$). Yet, Experiment 1 showed that motivation explained all of the variance shared between learning ability (as indexed by PA recall accuracy) and consistency (as indexed by proportions of off task thought), whereas Experiment 2 demonstrated that consistency actually explained unique variance in learning ability when accounting for motivation—regardless of whether pre, post, or mean motivation scores were used (see Appendix B for more details).

To better address the discrepant results, we pooled data across both experiments and reran the regression using post-motivation scores. Missing cases were excluded via pairwise deletion, leaving 296 participants with WM scores, 293 with LTM scores, and 302 participants with scores for PA recall accuracy, motivation, and proportion of off task thought. The simultaneous regression explained 40% of the variance in PA recall accuracy, $F(4, 281) = 47.35$, $p < .001$. Importantly, both post motivation ($\beta = .27$, $p < .001$) and proportions of off task thought ($\beta = -.14$, $p = .016$) accounted for unique variance in recall accuracy. Thus, the ability to consistently maintain attention on task was *partially* the result of decreased motivation. Individuals with higher motivation more consistently allocated attention to task-relevant information. In being more consistently focused on the task at hand, these individuals with high motivation were less susceptible to attentional capture by task-irrelevant sources, which contributed to better learning overall. But motivation does not seem to be the only important explanatory variable at play, given that there appear to be individuals with high levels of motivation who still struggle to keep their attention on task.

Unfortunately, we could not include our index of the intensity of attention, TEPRs, in the combined experimental analysis, since TEPRs were only assessed in Experiment 2. Nonetheless, Experiment 2 revealed that the best learners also tend to have larger pupillary responses at encoding, and, critically, individuals with larger pupil dilation tend to report (slightly) higher levels of motivation. So, individuals who allocated a greater proportion of their available attentional resources (more intensity) to the learning material tended to be more motivated and better learned the task. These results are consistent with the idea that participants mobilize effort to increase attention to the task in an attempt to boost overall task performance (e.g., Botvinick & Braver, 2015; Kanfer & Ackerman, 1989; Westbrook & Braver, 2015). Future work should aim to better disentangle how motivation and attentional factors influence one another, as well as other self-regulatory processes (e.g., performance monitoring, strategy use, decisions about effort expenditure), across the duration of the learning period.

Experiment 2 further revealed that intensity and consistency are related, insofar that TEPRs negatively correlated with lapses of attention during learning. This finding could mean that devoting fewer attentional resources to encoding in itself increases one's susceptibility to attentional lapses. Specifically, intensity is thought to be critically important for variation in overall task-engagement (Kahneman, 1973). Being in a state of lower task engagement on every trial (due to low intensity) could make it easier for off task thoughts (e.g., pressing concerns like a fight with one's spouse before work) to penetrate the focus of attention. Yet another possibility is that differences in intensity could be due to variation in consistency. Namely, when experiencing a lapse of attention, there is a temporary reduction in intensity to the current task. If an individual is off task more than they are on task, then their overall estimate of intensity would be lower than would be the case if they were predominantly on task.

Although changes in intensity appear to co-occur with changes in consistency, both aspects of attention do seem to have distinguishable effects on learning. That is, TEPRs and proportions of off task thought accounted for unique variance in associative learning when controlling for motivation, WM, and more general LTM abilities. Follow-up analyses suggested that one way in which these two aspects of attention are distinct is in terms of motivation. Note that higher motivation was associated with higher intensity, but the association was weak. A path analysis revealed that the slight increase in task engagement associated with increased motivation did not appear to be strong enough to actually impact learning—at least when accounting for variance shared with consistency. Indeed, variance partitioning showed that TEPRs (intensity) and motivation essentially shared no variance in predicting learning while controlling for shared variance with off task thought (consistency). Motivation, instead, primarily exerted its influence on learning ability by determining how consistently attention was allocated to various on vs off task activities during learning. In contrast to previous work (e.g., Botvinick & Braver, 2015; Westbrook & Braver, 2015), these results suggest that increased consistency may better explain the positive association between motivation and task performance than increased attentional effort (intensity)—at least within a learning and memory context.

Limitations and future directions

There are, however, a few caveats of the present study that are worth mentioning. First and foremost, we should be cautious in interpreting the strength of the effect of motivation on consistency (as indexed by reports of on vs off task thought) given that both were measured via self-report methods. These measures likely share some method variance. An additional issue worth noting concerns the timing of when these motivation measures are administered. Experiment 1 measured motivation only once (post-task), whereas Experiment 2 measured motivation twice (once pre-task and once post-task). When exclusively examining *post* motivation scores, the motivation-recall correlation in Experiment 1 ($r = .52$) was not significantly different from the correlation observed in Experiment 2 ($r = .35$), Fisher's $Z = 1.81$, $p = .07$. However, the motivation-recall correlation in Experiment 1 using post motivation was significantly larger than the motivation-recall correlations using *pre* motivation ($r = .21$, Fisher's $Z = 3.11$, $p = .002$) and *mean* motivation ($r = .32$, Fisher's $Z = 2.10$, $p = .036$) in Experiment 2. Accordingly, we suspect that correlations relying solely on post motivation measures are somewhat inflated by performance and awareness of attentional lapses to some degree (e.g., individuals who know they were performing inadequately may react to their poor performance by rating their motivation as lower). We advocate measuring motivation at least twice (pre and post) in a manner similar to what was done in Experiment 2. In doing so, changes in motivation across the task are still being captured if the dependent variable is mean motivation, and this variable is less influenced by said confounds.

As previously mentioned, motivation accounted for some learning-related differences in intensity and consistency. However, despite being highly motivated, some individuals were both unable to devote large amounts of attention to encoding and less able to consistently keep attention focused on task. Admittedly, we suspected that one's attentional resource capacity would partly explain variance unique to intensity as well as shared variance between intensity and consistency (independent of motivation). Yet no significant correlations were observed between TEPRs and WM in either experiment. That said, we don't think these null results can conclusively rule out the possible influence of attentional resource capacity. Note that the capacity of WM is typically defined as the number of items that participants can maintain in the current focus of attention (Cowan, 2001), meaning people with high WM capacity have more space (i.e., "slots") available for active maintenance and temporary storage of information (see Fukuda, Vogel, Mayr, & Awh, 2010). Based on this definition, we don't believe the capacity of WM is synonymous with the capacity of attentional resources. Even if one were to make such an assumption, research (Unsworth, Fukuda, Awh, & Vogel, 2014) has shown that there are multiple sources of variance *within* measures of WM. Thus, attention control or LTM retrieval abilities could obscure a relationship between attentional capacity (as measured by complex span tasks) and the intensity of attention (as indexed via TEPRs). Alternative measures of attentional resource capacity should continue to be explored.

Relatedly, since the ability to control attention in the presence of interference or other potent distractors is a theoretically important component of WM (Engle & Kane, 2004), we also expected WM to explain variation unique to consistency. Prior work has shown that lapses of attention/off task thoughts (inconsistency) negatively correlate with WM (Kane et al., 2016, 2017; McVay & Kane, 2010a, 2012; Mrazek et al., 2012; Robison & Unsworth, 2015, 2018, Unsworth & McMillan, 2013, 2014), but WM did not correlate with proportions of off task thought in either experiment. An important consideration here, though, is that previous research has primarily examined the relationship between WM and off task thoughts on attention control tasks (e.g., antisaccade, stroop, and flankers), and these correlations at the task level are also relatively weak (rs $= -.03$ to $-.17$; see Unsworth et al., in press). So, it is possible that both of the current studies were underpowered to detect such an effect (269 participants would be required to detect $r = -.17$ with 80% power).

Finally, intensity and consistency are likely not the only two aspects of attention abilities important for learning. For instance, where attention is selectively directed in the external environment may be another important factor. Indeed, Miller and Unsworth (2020) assessed the possible role of this selectivity component of attention by examining which information individuals attend to most during verbal associative learning (indexed by eye-fixations) and whether this is, in turn, related to recall performance. We reasoned that low learning individuals may perform worse because these individuals may selectively direct their attention towards only one of the relevant stimuli (e.g., the cue word) as opposed to both the target and cue word. However, across two experiments, learning was unrelated to where individuals allocated attention. Since we failed to detect any relations between learning and selective attentional focus, we did not incorporate those analyses in the present manuscript. Interestingly, though, when we retroactively computed the same variables described in Miller and Unsworth (2020) using our pupil data from Experiment 2, individuals who best learned the task in the present study appeared to fixate more on the target word ($r = .25$, $p < .001$, $N = 165$; see Appendix C for analyses with fixation variables). Future work clearly needs to further clarify which aspects of attention are important for learning and also determine how these attentional abilities are different and/or similar to differences in the ability to restrain, constrain, and sustain attention (among others). Accordingly, future research would also benefit from more powerful techniques that reduce measurement error, such as structural equation modeling, to better ascertain whether or not the intensity and consistency of attention are, in fact, reliability distinct constructs.

Conclusion

The present study is the first to provide evidence in favor of the notion that, while related constructs, intensity and consistency are not mere manifestations of the same general ability to control attention during learning. Rather, both aspects of attention appear to be partially distinct, important predictors of learning ability. The most successful learners seemingly allocate more attention to items during encoding (high intensity) and are also less susceptible to lapses of attention during this period (high consistency). Yet some individuals sufficiently learn to-be-remembered material despite having a particular deficit (e.g., high intensity + low consistency vs low intensity + high consistency). Furthermore, while the intensity and consistency of attention are both related to motivation and (to a lesser extent) general LTM abilities, motivation appears to be a factor more so implicated in the consistency of attention. Overall, the intensity and consistency of attention are likely separate, multifaceted constructs differentially influenced by a variety of factors. In order to understand learning and individual differences in learning ability, we undoubtedly need to further explore the complex nature of how attention is allocated during encoding and how these relations change as a function of numerous factors.

Credit authorship contribution statement

Ashley L. Miller: Conceptualization, Methodology, Software, Writing - original draft, Data curation, Investigation, Formal analysis, Visualization. **Nash Unsworth:** Conceptualization, Methodology, Software, Writing - review & editing, Supervision, Resources.

Appendix A

See Figs. A1 and A2.

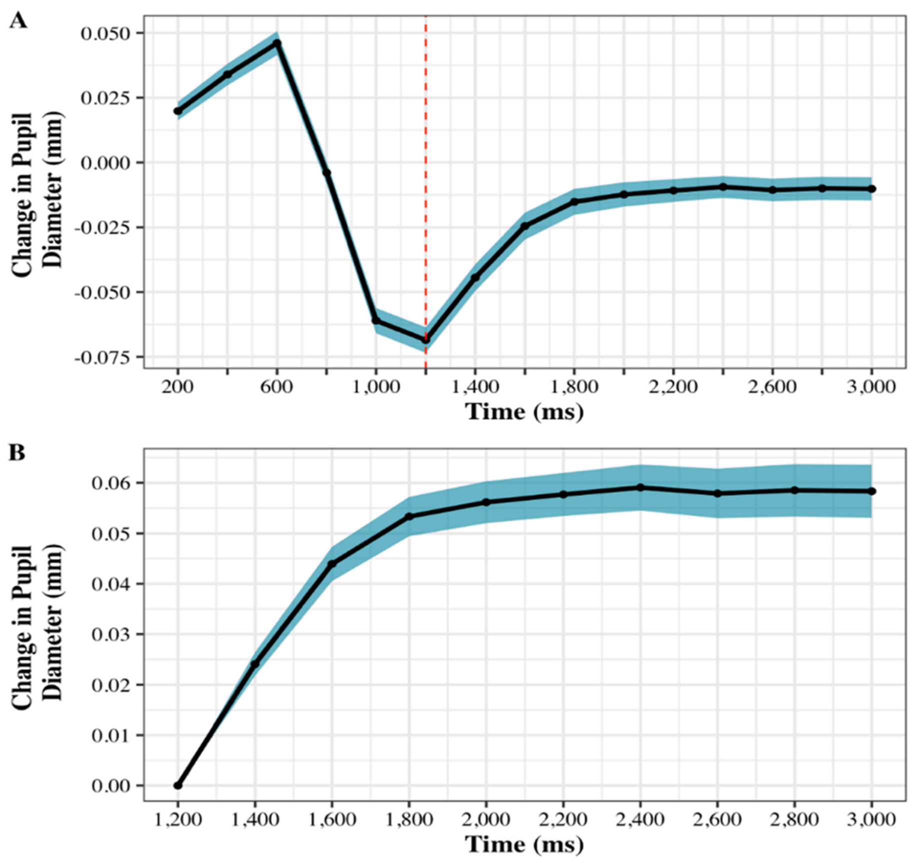

Fig. A1. (A) Change in pupil diameter (mm) across the entire 3-s encoding period for each word-pair. Vertical dashed line illustrates the end of the pupillary light reflex. (B) Change in pupil diameter during encoding using 1200 ms as our starting point and subtracting mean dilation at 1200 ms from each of the following bins. Shaded areas reflect 1 standard error of the mean.

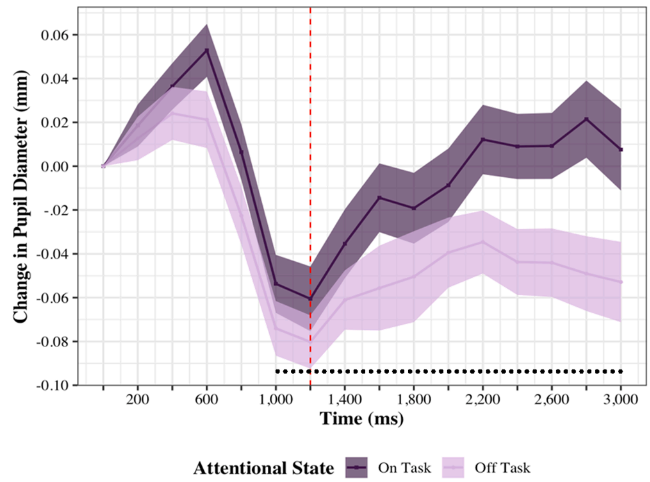

Fig. A2. Change in pupil diameter (mm) across the entire 3-s encoding period for each word-pair as a function of attentional state (on task vs off task). Vertical dashed line illustrates the end of the pupillary light reflex. Horizontal dashed line illustrates timepoints in which pairwise comparisons were significantly different when adjusting for multiple comparisons via the Bonferroni correction (p's $< .0033$).

Appendix B

Below we present the correlation matrix with motivation broken down by pre and post assessments. We then include the regression tables using pre and post motivation in isolation.

See Tables B1-B3.

Table B1
Correlations with pre and post motivation measures in Experiment 2.

Measure	1	2	3	4	5	6	7
1. PAacc	–						
2. WM	.12	–					
3. LTM	.48***	.25**	–				
4. PropOffTask	−.44***	−.01	−.16*	–			
5. PreMotivation	.21**	.01	.14	−.33***	–		
6. PostMotivation	.35***	−.04	.10	−.61***	.67***	–	
7. TEPR	.37***	−.10	.23**	−.27**	.10	.22**	–

Note. * $p<.05$, ** $p<.01$, *** $p<.001$; PAacc = paired associates cued recall accuracy, WM = working memory factor composite, LTM = long-term memory ability factor composite, PropOffTask = proportion of off task thoughts (i.e., attentional lapses), TEPR = mean task evoked pupillary response at the last encoding bin (the final 200 ms).

Table B2
Simultaneous regression predicting PA cued recall accuracy with pre motivation scores.

Variable	N	β	t	sr^2	R^2	F
LTM	164	.37	5.88***	.12		
WM	162	.04	.54	.001		
PropOffTask	165	−.32	−4.44***	.08		
PreMotivation	165	.03	.46	.001		
TEPR	148	.20	2.78**	.03	.40	18.52***

Note. * $p<.05$, ** $p<.01$, *** $p<.001$; Participants with missing data were excluded from this analysis via pairwise deletion.

Table B3
Simultaneous regression predicting PA cued recall accuracy with post motivation scores.

Variable	N	β	t	sr^2	R^2	F
LTM	164	.37	5.28***	.12		
WM	162	.04	.61	.002		
PropOffTask	165	−.26	−3.09**	.04		
PostMotivation	165	.12	1.43	.009		
TEPR	148	.19	2.69**	.03	.41	19.13***

Note. * $p<.05$, ** $p<.01$, *** $p<.001$; Participants with missing data were excluded from this analysis via pairwise deletion.

Appendix C

Fixations. Consistent with Miller and Unsworth (2020), four regions of interest (ROIs) corresponding to four locations (Cue, Target, Fixation, Other) were defined. The location *Fixation* simply represents the central fixation cross between the *Target* word and the *Cue* word, whereas *Other* refers to all other locations on the monitor. Proportions of fixations were determined by first summing the number of times an individual fixated on a given object during the 3-s encoding phase. Samples were collected every 16 ms. This number was then divided by an individual's total number of fixations. An inspection of the correlations in Table C1 suggests that individuals who best learned the task appeared to fixate more on the Target word ($r = .25$, $p < .001$, $N = 165$) and fixate less on the fixation cross between the word-pairs ($r = -.18$, $p = .02$, $N = 165$). Individuals who worse learned the task tended to show a slight tendency to fixate on other locations on the monitor ($r = -.16$, $p = .045$, $N = 164$). Importantly, adding mean proportion of Target fixations to the simultaneous regression reported in Experiment 2 did not alter the pattern of reported results. Due to shared variance with mean motivation ($r = .18$, $p = .018$), TEPRs ($r = .18$, $p = .028$), and proportions of off task thought ($r = -.19$, $p = .014$), the proportion of Target fixations did not significantly explain unique variance in learning ($\beta = .13$, $t = 1.95$, $p = .053$). As such, adding the mean proportion of Target fixations to the regression model did not significantly increase the amount of variance explained in learning, $\Delta R^2 = .016$, $F(1, 139) = 3.80$, $p = .053$

Switches. For each second of the encoding period, we summed Cue-to-Target and Target-to-Cue switches. A switch was defined as a fixation on either the Cue or Target ROI when the previously fixated ROI was the opposite. The mean was obtained by averaging across Cue-to-Target and Target-to-Cue switches. Consistent with Experiment 1 in Miller and Unsworth (2020), more switches overall during encoding were associated with better PA recall accuracy ($r = .21$, $p < .01$; see also Kamp & Zimmer, 2015).

Table C1
Correlations with all fixation variables in Experiment 2.

Measure	1	2	3	4	5	6	7	8	9	10
1. PAacc	–									
2. WM	.12	–								
3. LTM	.48***	.25**	–							
4. OffTask	−.44***	−.01	−.16*	–						
5. MeanMotiv	.32***	−.02	.13	−.53***	–					
6. TEPR	.37***	−.10	.23**	−.27**	.18*	–				
7. CueFix	.01	−.06	−.03	.07	−.08	.17*	–			
8. TargetFix	.25**	.10	.06	−.19*	.18*	.18*	−.18*	–		
9. FixFix	−.18*	−.04	.00	.05	−.07	−.20*	−.62***	−.60***	–	
10. Switches	.21**	.09	.10	−.14	.16*	.11	.34***	.09	−.23**	–

Note. * $p<.05$, ** $p<.01$, *** $p<.001$; Participants with missing data were excluded from this analysis via pairwise deletion. PAacc = paired associates cued recall accuracy, WM = working memory factor composite, LTM = long-term memory ability factor composite, OffTask = proportion of off task thoughts (i.e., attentional lapses), TEPR = mean task evoked pupillary response at the last encoding bin (the final 200 ms), CueFix = proportion of fixations on Cue word; TargetFix = proportion of fixations on Target word; FixFix = proportion of fixations on central Fixation cross; Switches = mean switches (gaze alterations) between the Target and Cue word.

References

Ackerman, P. L. (2011). 100 years without resting. In P. L. Ackerman (Ed.), *Cognitive fatigue: Multidisciplinary perspectives on current research and future applications* (pp. 11–43). American Psychological Association.

Anderson, N. D., Craik, F. I. M., & Naveh-Benjamin, M. (1998). The attentional demands of encoding and retrieval in younger and older adults: 1. Evidence from divided attention costs. *Psychology and Aging, 13*, 405–423.

Ariel, R., & Castel, A. D. (2014). Eyes wide open: Enhanced pupil dilation when selectively studying important information. *Experimental Brain Research, 232*, 337–344.

Baddeley, A., Lewis, V., Eldridge, M., & Thomson, N. (1984). Attention and retrieval from long-term memory. Journal of Experimental Psychology: General, 113, 518–540.

Bailey, H., Dunlosky, J., & Kane, M. J. (2008). Why does working memory span predict complex cognition? Testing the strategy affordance hypothesis. *Memory & Cognition, 36*, 1383–1390.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (pp. 142–162). New York: Cambridge University Press.

Binda, P., Pereverzeva, M., & Murray, S. O. (2013). Attention to bright surfaces enhances the pupillary light reflex. *Journal of Neuroscience, 33*, 2199–2204.

Botvinick, M., & Braver, T. (2015). Motivation and cognitive control: From behavior to neural mechanism. *Annual Review of Psychology, 66*, 83–113.

Chuah, Y. L., & Maybery, M. T. (1999). Verbal and spatial short-term memory: Common sources of developmental change? *Journal of Experimental Child Psychology, 73*, 7–44.

Conway, A. R. A., & Engle, R. W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General, 123*, 354–373.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–114.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51*, 42–100.

Craik, F. I. M., & Byrd, M. (1982). Aging and cognitive deficits: The role of attentional resources. In F. I. M. Craik, & S. E. Trehub (Eds.), *Aging and cognitive processes* (pp. 191–211). New York, NY: Plenum Press.

deBettencourt, M. T., Norman, K. A., & Turk-Browne, N. B. (2018). Forgetting from lapses of sustained attention. *Psychonomic Bulletin & Review, 25*, 605–611.

Engle, R. W. (1975). Pupillary measurement and release from proactive inhibition. *Perceptual and Motor Skills, 41*, 835–842.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 145–199). NY: Elsevier.

Friendly, M., Franklin, P. E., Hoffman, D., & Rubin, D. C. (1982). The Toronto word pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. *Behavioral Research Methods & Instrumentation, 14*, 375–399.

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review, 17*, 673–679.

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science, 2*, 156–168.

Garlitch, S. M., & Wahlheim, C. N. (2020). The role of attentional fluctuation during study in recollecting episodic changes at test. *Memory & Cognition, 48*, 800–814.

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74–78.

Goldinger, S. D., & Papesh, M. H. (2012). Pupil dilation reflects the creation and retrieval of memories. *Current Directions in Psychological Science, 21*, 90–95.

Granholm, E., & Steinhauer, S. R. (2004). Pupillometric measures of cognitive and emotional processes. *International Journal of Psychophysiology, 52*, 1–6.

Gross, M. P., & Dobbins, I. G. (2021). Pupil dilation during memory encoding reflects time pressure rather than depth of processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Heaver, B., & Hutton, S. B. (2011). Keeping an eye on the truth? Pupil size changes associated with recognition memory. *Memory, 19*, 398–405.

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem solving. *Science, 143*, 1190–1192.

Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology, 45*, 73–93.

Hockey, G. R. J. (2013). *The Psychology of Fatigue: Work, Effort, and Control*. New York, NY: Cambridge University Press.

Hutchison, K. A., Moffitt, C. C., Hart, K., Hood, A. V. B., Watson, J. M., & Marchak, F. M. (2020). Measuring task set preparation versus mind wandering using pupillometry. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*, 280–295.

Janisse, M. P. (1977). *Pupillometry: The psychology of the pupillary response*. Hemisphere Pub.

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology, 47*, 310–339.

Kafkas, A. (2021). Encoding-linked pupil response is modulated by expected and unexpected novelty: Implications for memory formation and neurotransmission. *Neurobiology of Learning and Memory, 107412*.

Kafkas, A., & Montaldi, D. (2011). Recognition memory strength is predicted by pupillary responses at encoding while fixation patterns distinguish recollection from familiarity. *The Quarterly Journal of Experimental Psychology, 64*, 1971–1989.

Kahneman, D. (1973). *Attention and effort*. Upper Saddle River, NJ: Prentice Hall.

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

Kalsbeek, J. W. H. (1968). Measurement of mental workload and of acceptable load: Possible applications in industry. *International Journal of Production Research, 7*, 33–45.

Kamp, S., & Zimmer, H. D. (2015). Contributions of attention and elaboration to associative encoding in young and older adults. *Neuropsychologia, 75*, 252–264.

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

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual differences perspective. *Psychonomic Bulletin & Review, 9*, 637–671.

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

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, 1017-1048.

Kanfer, R. (1987). Task-specific motivation: An integrative approach to issues of measurement, mechanisms, processes, and determinants. *Journal of Social and Clinical Psychology, 5*, 237–264.

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology, 74*, 657–690.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting & task performance*. Prentice-Hall Inc.

Maillet, D., & Rajah, M. N. (2013). Age-related changes in frequency of mind-wandering and task-related interferences during memory encoding and their impact on retrieval. *Memory, 21*, 818–831.

Maillet, D., & Rajah, M. N. (2014). Dissociable roles of default-mode regions during episodic encoding. *NeuroImage, 89*, 244–255.

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

McVay, J. C., & Kane, M. J. (2010a). Adrift in the stream of thought: The effects of mind wandering on executive control and working memory. In A. Gruszka, G. Matthews, & B. Szymura (Eds.), *Handbook of Individual Differences in Cognition. The Springer Series on Human Exceptionality*. New York, NY: Springer.

McVay, J. C., & Kane, M. J. (2010b). Does mind wandering reflect executive function or executive failure? Comment on Smallwood and Schooler (2006) and Watkins (2008). *Psychological Bulletin, 136*, 188–197.

McVay, J. C., & Kane, M. J. (2012). Why does working memory predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General, 141*, 302–320.

McVay, J. C., Kane, M. J., & Kwapil, T. R. (2009). Tracking the train of thought from the laboratory into everyday life: An experience-sampling study of mind-wandering in controlled and ecological contexts. *Psychonomic Bulletin & Review, 16*, 857–863.

Metcalfe, J., & Xu, J. (2016). People mind wander more during massed than spaced inductive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 978–984.

Miller, A. L., Gross, M. P., & Unsworth, N. (2019). Individual differences in working memory and long-term memory: The influence of intensity of attention to items at encoding as measured by pupil dilation. *Journal of Memory and Language, 104*, 25–42.

Miller, A. L., & Unsworth, N. (2020). Variation in attention at encoding: Insights from pupillometry and eye gaze fixations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*, 2277–2294.

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, 144*, 788–798.

Otero, S. C., Weekes, B. S., & Hutton, S. B. (2011). Pupil size changes during recognition memory. *Psychophysiology, 48*, 1346–1353.

Papesh, M. H., Goldinger, S. D., & Hout, M. C. (2012). Memory strength and specificity revealed by pupillometry. *International Journal of Psychophysiology, 83*, 56–64.

Peavler, W. S. (1974). Pupil size, information overload, and performance differences. *Psychophysiology, 11*, 559–566.

Porter, G., Troscianko, T., & Gilchrist, I. D. (2007). Effort during visual search and counting: Insights from pupillometry. *Quarterly Journal of Experimental Psychology, 60*, 211–229.

Redick, T. S., Shipstead, Z., Meier, M. E., Montroy, J. J., Hicks, K. L., Unsworth, N., … Engle, R. W. (2016). Cognitive predictors of a common multitasking ability: Contributions from working memory, attention control, and fluid intelligence. *Journal of Experimental Psychology: General, 145*, 1473–1492.

Robison, M. K., Miller, A. L., & Unsworth, N. (2020). A multi-faceted approach to understanding individual differences in mind-wandering. *Cognition, 198*, Article 104078.

Robison, M. K., & Unsworth, N. (2015). Working memory offers resistance to mind-wandering and external distraction in a context specific manner. *Applied Cognitive Psychology, 29*, 680–690.

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*, 85–98.

Rohrer, D., & Wixted, J. T. (1994). An analysis of latency and interresponse time in free recall. *Memory & Cognition, 22*, 511–524.

Schmidtke, H. (1976). Vigilance. In E. Simpson, & P. C. Weiser (Eds.), *Psychological and physiological correlates of work and fatigue* (pp. 193–219). Springfield, IL: Charles C Thomas.

Seibert, P. S., & Ellis, H. C. (1991). Irrelevant thoughts, emotional mood states, and cognitive task performance. *Memory & Cognition, 19*, 507–513.

Seli, P., Cheyne, J. A., Xu, M., Purdon, C., & Smilek, D. (2015). Motivation and mind wandering: Implications for assessments of task-unrelated thought. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 41*, 1417–1425.

Seli, P., Schacter, D. L., Risko, E. F., & Smilek, D. (2019). Increasing participant motivation reduces rates of intentional and unintentional mind wandering. *Psychological Research, 83*, 1057–1069.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron, 79*, 217–240.

Smallwood, J., Baracaia, S. F., Lowe, M., & Obonsawin, M. (2003). Task unrelated thought whilst encoding information. *Consciousness and Cognition, 12*, 452–484.

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin, 132*, 946–958.

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*, 370–381.

Taikh, A. (2014). *Effects of Levels-of-processing and Test-list Context on Recognition and Pupil Dilation [Unpublished master's thesis]*. Calgary, AB: University of Calgary.

Thomson, D. R., Seli, P., Besner, D., & Smilek, D. (2014). On the link between mind wandering and task performance over time. *Consciousness and Cognition, 27*, 14–26.

Thomson, D. R., Smilek, D., & Besner, D. (2014). On the asymmetric effects of mind-wandering on levels of processing at encoding and retrieval. *Psychonomic Bulletin and Review, 21*, 728–733.

Unsworth, N. (2015). Consistency of attentional control as an important cognitive trait: A latent variable analysis. Intelligence, 49, 110-128.

Unsworth, N. (2019). Individual differences in long-term memory. *Psychological Bulletin, 145*, 79–139.

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory. *Cognitive Psychology, 71*, 1–26.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*, 498–505.

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

Unsworth, N., & McMillan, B. D. (2014). 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.

Unsworth, N., & McMillan, B. D. (2017). Attentional disengagements in educational contexts: A diary investigation of everyday mind-wandering and distraction. *Cognitive Research: Principles and Implications, 2*, 32.

Unsworth, N., McMillan, B. D., Brewer, G. A., & Spillers, G. J. (2012). Everyday attention failures: An individual differences investigation. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 38*, 1765–1772.

Unsworth, N., & Miller, A. L. (2021). Encoding dynamics in free recall: Examining attention allocation with pupillometry. *Memory & Cognition, 49*, 90–111.

Unsworth, N., Miller, A. L., & Robison, M. K. (2020). Individual differences in lapses of sustained attention: Ocolumetric indicators of intrinsic alertness. *Journal of Experimental Psychology: Human Perception and Performance, 46*, 569–592.

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent variable analysis of the relationship between processing and storage. *Memory, 17*, 635–654.

Unsworth, N., & Robison, M. K. (2015). Individual differences in the allocation of attention to items in working memory: Evidence from pupillometry. *Psychonomic Bulletin & Review, 22*, 757–865.

Unsworth, N., & Robison, M. K. (2017a). A locus coeruleus-norepinephrine account of individual differences in working memory and attention control. *Psychonomic Bulletin & Review, 24*, 1282–1311.

Unsworth, N., & Robison, M. K. (2017b). The importance of arousal for variation in working memory and attention control: A latent variable pupillometry study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*, 1962–1987.

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

Unsworth, N., & Robison, M. K. (2020). Working memory and sustained attention: A cognitive-energetic perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*, 77–103.

Unsworth, N., Robison, M.K., & Miller, A.L. (in press). Individual differences in lapses of attention: A latent variable analysis. Journal of Experimental Psychology: General.

Wang, C., Brien, D. C., & Munoz, D. P. (2015). Pupil size reveals preparatory processes in the generation of pro-saccades and anti-saccades. *European Journal of Neuroscience, 41*, 1102–1110.

Ward, A. F., & Wegner, D. M. (2013). Mind-blanking: When the mind goes away. *Frontiers in Psychology, 4*, 650.

Westbrook, A., & Braver, T. S. (2015). Cognitive effort: A neuroeconomic approach. *Cognitive, Affective & Behavioral Neuroscience, 15*, 395–415.

Wetzel, N., Einhäuser, W., & Widmann, A. (2020). Picture-evoked changes in pupil size predict learning success in children. *Journal of Experimental Child Psychology, 192*, Article 104787.

Xu, J., & Metcalfe, J. (2016). Studying in the region of proximal learning reduces mind wandering. *Memory & Cognition, 44*, 681–695.