Fuel for Working Memory:

The Same Resource for Consolidation and Refreshing?

university of groningen

faculty of science and engineering

Maximilian Linde

(s2512491)

January 2018

Minor Thesis MSc Program of Behavioural and Cognitive Neurosciences

Faculty of Science and Engineering

University of Groningen

Supervised by: dr. Mark. R. Nieuwenstein

Second evaluator: dr. Sebastiaan Mathôt

Abstract

Working memory (WM) allows us to temporarily hold on to and manipulate information for ongoing cognition. WM consolidation is the process of transferring sensory information into WM. Refreshing is a recently proposed maintenance mechanism that is assumed to prevent forgetting by bringing the contents of WM into the focus of attention. Refreshing an item increases its accessibility, observable in reduced reaction times. Evidently, both consolidation and refreshing require attention. However, there is a lack of research on the interplay between consolidation and refreshing. The overall goal of the current investigation was to test whether consolidation interferes with people's ability to refresh WM representations. In the first experiment, we demonstrated the feasibility of a simplified version of the instructed-refreshing paradigm by Vergauwe and Langerock (2017). This methodological verification was crucial for the implementation of this paradigm as part of the dual-task paradigm in the second experiment. In the second experiment, we investigated whether consolidation and refreshing utilize the same or separate resources. Critically, we manipulated whether refreshing had to be done during or after consolidation of a picture. A decreased refresh benefit should be observed during consolidation, compared to after consolidation, if both processes require the same resource, whereas no performance decrement should be evident if separate resources are used. Unfortunately, we could not properly probe the two competing hypotheses because no comparison of refresh benefits during and after consolidation was legitimate, since we did not observe a refresh benefit after consolidation. Nevertheless, we provide suggestions for methodological improvements that could be seized in future experiments.

Keywords: working memory, refreshing, consolidation, attention, attentional blink

Table of Contents

Fuel for Working Memory:
The Same Resource for Consolidation and Refreshing?

Everyday life is filled with plenty of situations that require us to hold on to certain information, which we then process or manipulate in accordance with our current cognitive and behavioral needs. Oftentimes, this duality of maintenance and manipulation proceeds in an alternating manner; that is, the output of a manipulation of maintained information can serve as the new target of maintenance, which can then be further processes, and so forth. To illustrate, let us pretend that we do some mental arithmetic and wish to divide the product of seven and eight by four. Obviously, we need to keep all three numbers and the mathematical operators in mind. Then, we need to multiply seven and eight and store the result of 56. Subsequently, we need to remember this result together with the divisor of four and, lastly, calculate the end result of 14. Although this operation seems very simple as a whole, we have to acknowledge that even such a mundane task compels us to switch between maintaining and manipulating the relevant information to reach our current goal. Examples of tasks of a similar nature are abundant. For instance, engaging in conversation demands that we maintain and process what the other person said, in order to come up with response ideas that we combine and, eventually, formulate and express; or, alternatively, if we want to enter an address into the car navigation system, we need to keep the address active in our mind until we arrive at the car (Olivers, Peters, Houtkamp, & Roelfsema, 2011). The underlying component of the human cognitive architecture that allows us to master all of these operations is working memory (WM).

WM is a limited-capacity system that temporarily holds on to and manipulates information for current cognition and action, forming an interface between perception, long-term memory systems, and behavior (e.g., Baddeley & Hitch, 1974; Baddeley, 2003a; Oberauer, 2002; Vergauwe & Langerock, 2017). In essence, WM is a state that encompasses those information or memory traces that are activated and that are currently thought about or worked upon. Emphasizing the key role of WM in the human cognitive architecture, Haberlandt (1997) titled it as the “hub of cognition” (p. 212) and Goldman-Rakic (1992) revered it as “perhaps the most significant achievement of human mental evolution” (p. 111). There is a wealth of literature attesting to the central importance of WM, showing that its capacity predicts and contributes to, among other things, fluid intelligence, academic achievement, reasoning and problem solving abilities, learning, language comprehension, and mental arithmetic (e.g., Barrouillet, 1996; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Daneman & Carpenter, 1980; Daneman & Merikle, 1996; DeStefano & LeFevre, 2004; Engle, Kane, & Tuholski, 1999; Engle, Cantor, & Carullo, 1992; Engle, Tuholski, Laughlin, & Conway, 1999; Halford, Wilson, & Phillips, 1998; Harrison, Shipstead, & Engle, 2015; Kane et al., 2004; Kyllonen & Christal, 1990; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002). In line with that, it was demonstrated that people with a larger WM capacity are less inclined to have their minds wander when performing a task, compared to people with a lower WM capacity (Kane et al., 2007). Clearly, then, WM is a central component of human cognition with wide-ranging applications and implications in everyday life. Notwithstanding, WM functioning is not without its boundaries.

In contrast to long-term memory, which can store an almost infinite amount of memories for up to an entire life time (Brady, Konkle, Alvarez, & Oliva, 2008; Standing, 1973), WM is restricted in both its capacity and retention time (e.g., Barrouillet, Uittenhove, Lucidi, & Langerock, 2017). Initially, the average capacity of WM was estimated to be approximately seven chunks of information (Miller, 1956), which was later updated and adjusted to be around four (e.g., Cowan, 2001; Cowan, 2010). The difference between those two estimates is that the former relies solely on storage processes and allows participants to chunk (i.e., combine) information, whereas the latter probes the storage capacity of WM while people process other material at the same time. Arguably, the latter is better suited as an estimate of WM capacity because it more closely captures the nature of WM in terms of its two functions of maintaining and processing or manipulating information (Baddeley & Hitch, 1974; Vergauwe, Camos, & Barrouillet, 2014). It should be acknowledged, however, that these estimates can vary depending on the circumstances of the testing situation (Cowan, 2010). Furthermore, in contrast to long-term memory, the time scale of retention in WM is limited to only a few seconds (Barrouillet et al., 2017). The contents of WM are very fragile; they are simply forgotten within seconds if they are not rehearsed in some way (e.g., Baddeley, 2003a; Barrouillet & Camos, 2012; Brown, 1958; Peterson & Peterson, 1959).

Crucially, WM involves several processes with different functions. Probably the three most essential ones, currently receiving a lot of attention by researchers, are consolidation, refreshing, and removal (Ricker, Nieuwenstein, Bayliss, & Barrouillet, 2018). Consolidation refers to the process of transferring transient perceptual information into relatively durable WM representations (e.g., Chun & Potter, 1995; Jolicœur & Dell'Acqua, 1998; Vogel & Luck, 2002). Refreshing is one of several proposed maintenance mechanisms of WM (e.g., Camos & Barrouillet, 2014; Raye, Johnson, Mitchell, Greene, & Johnson, 2007; Souza, Rerko, & Oberauer, 2015; Vergauwe & Langerock, 2017), with the goal of keeping information in WM active. Lastly, removal constitutes a process whereby items in WM are eliminated for the purpose of updating the contents of WM or avoiding interference from distracting items (Ecker, Lewandowsky, & Oberauer, 2014; Oberauer, 2001; Oberauer & Lewandowsky, 2016; Souza, Rerko, & Oberauer, 2014). Removal is mostly relevant for interference-based accounts of forgetting in WM (e.g., Oberauer, 2001; Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012; Oberauer & Lewandowsky, 2014; Lewandowsky, Oberauer, & Brown, 2009; see also Barrouillet et al., 2017, for a recent review). A discussion of removal processes is beyond the scope of the present study and, thus, will not be further discussed.

Working Memory Consolidation

WM consolidation constitutes the transfer of transient perceptual information into WM (e.g., Chun & Potter, 1995; Jolicœur & Dell'Acqua, 1998; Vogel & Luck, 2002) and is typically studied with the attentional blink (AB) paradigm (Raymond, Shapiro, & Arnell, 1992; see also Chun & Potter, 1995; Nieuwenstein, Chun, van der Lubbe, & Hooge, 2005; Vogel, Luck, & Shapiro, 1998). There are several variants of the AB paradigm. The most common version uses a very fast stream of visual stimuli (approximately at a rate of 10 Hz), called rapid serial visual presentation (RSVP), whereby two visual targets (i.e., first target [T1] and second target [T2], accordingly) are embedded within multiple distractors (e.g., Broadbent & Broadbent, 1987; Chun & Potter, 1995; Kanwisher, 1987; Weichselgartner & Sperling, 1987). Targets can be distinguished from distractors by one of several possible characteristics, for instance, color, rotation, category (e.g., digit vs. letter), and so forth. Subjects are asked to report the identity of T1 and T2 at the end of each trial (e.g., Broadbent & Broadbent, 1987; Raymond et al., 1992). Crucially, the time interval between the onsets of the two targets, called the stimulus onset asynchrony (SOA), is systematically varied (Di Lollo, Kawahara, Ghorashi, & Enns, 2005). What is typically found is that people have a marked difficulty reporting T2 when it follows T1 within less than approximately 500 ms (e.g., Chun & Potter, 1995; Raymond et al., 1992). In analogy to the overt blink of the eyes, this identification deficit of T2 was termed the attentional blink or AB (e.g., Chun & Potter, 1995; Dux & Marois, 2009; Raymond et al., 1992; Vogel et al., 1998).

There are a variety of accounts that try to explain the AB phenomenon. In a seminal paper, Chun and Potter (1995) argued that the AB arises due to the duration of the capacity-limited process of encoding T1 into WM (see also Jolicœur & Dell'Acqua, 1998). More specifically, this view proposes that the T2 identification deficit stems from a serial bottleneck; as long as central capacities are deployed for the consolidation of T1, the consolidation of T2 is postponed, which results in a deficiency to identify and report T2. According to this account, T2 performance should increase as SOA increases because with increasing time between the two targets, it is more likely that central capacities are available again for processing T2.

This sheer limited-capacity account was queried by a series of studies. In contrast to the predictions of this account, T2 identification accuracy does not always increase monotonically with SOA. In particular, if T2 follows T1 within less than 100 ms (e.g., Potter, Chun, Banks, & Muckenhoupt, 1998; see also VanRullen & Koch, 2003) and without any distractors in between, T2 performance is usually quite high, a phenomenon called lag 1 sparing (e.g., Hommel & Akyürek, 2005; Potter et al., 1998; Visser, Bischof, & Di Lollo, 1999). The AB is even mitigated when multiple targets are presented, which is called successive target advantage or spreading the sparing (Di Lollo et al., 2005; Kawahara, Enns, & Di Lollo, 2006; Nieuwenstein & Potter, 2006; Nieuwenstein, 2006; Olivers, Van der Stigchel, & Hulleman, 2007). It is worth mentioning, however, that sparing is sometimes associated with several costs (see Wyble, Bowman, & Nieuwenstein, 2009), namely increased repetition blindness (Chun, 1997b; Kanwisher, 1987), more order errors (Chun & Potter, 1995), and more conjunction errors (Chun, 1997a). Obviously, the limited-capacity account does not readily accommodate these findings since it states that the time that has elapsed since the onset of T1 determines whether central capacities are available again for the consolidation of T2.

Nieuwenstein and colleagues (2009) differentiated between two classes of accounts that can potentially explain these sparing effects, namely those based on distractor interference and those based on delayed attentional reengagement. The former posits that the

AB arises due to interference from distractors presented amidst the two targets, hampering the processing of T2 (Di Lollo et al., 2005; Olivers & Nieuwenhuis, 2006; Olivers et al., 2007). In the studies on sparing effects (viz. Di Lollo et al., 2005; Hommel & Akyürek, 2005; Kawahara et al., 2006; Nieuwenstein & Potter, 2006; Nieuwenstein, 2006; Olivers et al., 2007; Potter et al., 1998; Visser et al., 1999) no distractors were presented. Hence, no interference could have been involved and, thus, the report of the targets was quite accurate.

Delayed-reengagement accounts (e.g., Nieuwenstein et al., 2005; Nieuwenstein & Potter, 2006; Nieuwenstein, 2006; Weichselgartner & Sperling, 1987), in contrast, assert that the AB stems from a postponement of reengaging attention upon the arrival of T2. More specifically, the detection of T1 deploys attention that amplifies further processing of T1. However, as soon as the consolidation of T1 is initiated, attention is suppressed for the duration of the consolidation process. As a result, if the trailing target (i.e., T2) falls within the attentional suppression window, it is less able to reengage attention in time and, thus, the consolidation of this target fails or is at least delayed. Yet, this also implies that T2 should be processed and consolidated if it falls within the window of attentional deployment. In line with that, the delayed-reengagement account suggests that sparing occurs because targets fall within the transient attentional deployment window of the previous target, with each target maintaining the deployment of attention. Thus, the presentation of consecutive targets evokes a form of sustained attention, resulting in participants being able to report these targets (Nieuwenstein et al., 2009; see also Wyble et al., 2009). Further evidence for this claim comes from studies in which T2 was precued (Kawahara et al., 2006; Nieuwenstein et al., 2005; Nieuwenstein, 2006; Nieuwenstein et al., 2009; Olivers et al., 2007; Vul, Nieuwenstein, & Kanwisher, 2008; Wyble et al., 2009; Wyble, Potter, Bowman, & Nieuwenstein, 2011; Zhang, Shao, Nieuwenstein, & Zhou, 2008). In particular, when T2 is preceded by another target (i.e., a cue) the AB is vastly attenuated. According to the delayed-reengagement account, the cue triggers the reengagement of attention, having the consequence that T2 falls within the delayed transient attentional window elicited by the cue (Nieuwenstein et al., 2009). Nieuwenstein and colleagues (2005) showed that the cue reaches its highest potential in rescuing T2 identification if it is presented approximately 100-200 ms in advance of T2. This finding provides an estimate of the delay of reengaging attention during the consolidation of a target.

Importantly, Nieuwenstein, Potter, and Theeuwes (2009) presented compelling evidence against the credibility of distractor-interference accounts. In their study, two visual targets were embedded within a sequence of distractor stimuli. Crucially, they manipulated whether inter-target distractors were presented or not. Interestingly, an AB was observed despite the lack of inter-target distractors, provided that T2 was masked and the T2 task was difficult enough (e.g., by reducing the conventional presentation time of T2 from 100 to 58.3 ms). Even beyond that, it was shown that the distractors preceding T1 are not necessary for the occurrence of an AB (Nieuwenstein, Van der Burg, Theeuwes, Wyble, & Potter, 2009). Thus, the presentation of T1 and a masked T2 is enough to elicit a reliable AB; this arrangement of stimuli is called the skeletal RSVP or skeletal AB paradigm (e.g., Nieuwenstein et al., 2009; Ward, Duncan, & Shapiro, 1997). The most essential utility of the skeletal AB paradigm is a narrowed scope of possible explanations of the AB (Nieuwenstein et al., 2009). That is to say, since the skeletal RSVP paradigm does not entail any distractors, the T2 identification deficit could not possibly stem from distractor interference (viz. Di Lollo et al., 2005; Olivers & Nieuwenhuis, 2006; Olivers et al., 2007). Any debilities to report T2 can only be ascribed to the processing of T1 (Nieuwenstein et al., 2009). For these reasons and because of its simplicity, we employed a modified version of the skeletal AB paradigm in the second experiment.

The episodic simultaneous type, serial token (eSTST) model (Wyble et al., 2009; see also Bowman & Wyble, 2007) is a computational model (i.e., a neural network) that aims to explain the process of WM consolidation and the AB phenomenon more generally. Indeed, the eSTST model incorporates various findings from AB studies (e.g., lag 1 sparing and effects of precuing T2) and reproduces AB data quite accurately (Wyble et al., 2009). The critical elements of this model are the input nodes, type and token nodes (Kanwisher, 1987), a blaster, and a binding pool. Upon the arrival of a target stimulus, input nodes receive excitatory signals from early visual areas. If stimuli do not match the task demands (e.g., distractors), input nodes are inhibited in order to prevent further processing. However, if targets are presented, the activated input nodes excite the blaster, which dynamically controls the deployment of attention. The blaster, in turn, provides transient attentional resources for the input nodes, amplifying their activation. Once the input nodes have sufficient activation, they excite type nodes. Types constitute a workspace in which visual input is analyzed, and features and the identity of the stimulus are extracted. Importantly, type nodes do not entail any episodic information, for example the temporal context of stimuli. With the support of attention, types are then bound to tokens in the binding pool. Tokens are WM representations that are accompanied with episodic information. Crucially, once this tokenization process is initiated, the blaster is inhibited and, thus, is not able to amplify activation in the input nodes. As a result, during this binding or tokenization period, new incoming stimuli do not receive amplification from the blaster and they are less likely to reengage attention. Thus, the processing of new incoming stimuli is suppressed, which results in the T2 identification deficit that constitutes the AB. The binding process is assumed to represent WM consolidation. Obviously, then, the blaster is competitively regulated, with tokens inhibiting and input nodes exciting the blaster. Interestingly, when consecutive targets are presented (Di Lollo et al., 2005; Kawahara et al., 2006; Nieuwenstein & Potter, 2006; Nieuwenstein, 2006;

Olivers et al., 2007) or T2 is precued (Kawahara et al., 2006; Nieuwenstein et al., 2005; Nieuwenstein, 2006; Nieuwenstein et al., 2009; Olivers et al., 2007; Vul et al., 2008; Wyble et al., 2009; Wyble et al., 2011; Zhang et al., 2008), the competition is biased towards excitation from the input nodes. Consequently, attentional deployment by the blaster is sustained, resulting in the consolidation of multiple targets and, ultimately, accurate identifications of those targets.

It is well known that the magnitude and the duration of the AB can vary dramatically depending on the nature of the stimuli used for T1. For example, an AB is only observed if people are aware and, thus, can report the identity of T1 (Nieuwenstein et al., 2009). Furthermore, if T1 is a picture of negative emotional valence, the AB is increased compared to a neutral picture (Most, Chun, Widders, & Zald, 2005), probably because attention dwells longer on the T1 picture if it is emotionally arousing (e.g., Duncan, Ward, & Shapiro, 1994). Similarly, erotic images seem to elicit an emotion-induced blindness (Most, Smith, Cooter, Levy, & Zald, 2007). The authors recently found that memorable pictures (Bylinskii, Isola, Bainbridge, Torralba, & Oliva, 2015; Isola, Parikh, Torralba, & Oliva, 2011; Isola, Xiao, Torralba, & Oliva, 2011; Isola, Xiao, Parikh, Torralba, & Oliva, 2014; Xiao, Hays, Ehinger, Oliva, & Torralba, 2010) that need to be memorized and later recognized elicit a strong AB (Linde, Nieuwenstein, & van Rijn, 2016). Crucial for the current investigation, it was demonstrated that pictures of objects (Brady et al., 2008), which needed to be memorized in detail and had to be discriminated from the same object in a slightly different state at the end of the trial, elicit an enormous AB, both in duration and magnitude (Nieuwenstein & Kromm, 2017). The AB, however, was reduced when participants only had to memorize the gist of the pictures. Overall, these studies provide compelling evidence that pictures as T1 elicit a much stronger AB than simple alphanumeric characters, used most often in AB studies (e.g., Chun & Potter, 1995; Di Lollo et al., 2005; Nieuwenstein, 2006; Nieuwenstein et al., 2009;

Raymond et al., 1992). Because of the desirable property of state-pair pictures (Brady et al., 2008), if memorized in detail, to elicit a long-lasting AB, we decided to use the same stimuli as T1 in our second experiment.

Working Memory Refreshing

Representations in WM are fragile and forgotten over a few seconds (e.g., Baddeley, 2003a; Barrouillet & Camos, 2012; Barrouillet et al., 2017; Brown, 1958; Peterson & Peterson, 1959). There are several accounts that argue for distinct underlying causes of forgetting in WM. The two most prominent ones are based on interference (e.g., Lewandowsky, Duncan, & Brown, 2004; Lewandowsky et al., 2009; see also Oberauer & Kliegl, 2006; Oberauer & Lewandowsky, 2008; Oberauer et al., 2012, for excellent computational models based on interference) or, alternatively, temporal decay (e.g., Barrouillet, Bernardin, & Camos, 2004; Barrouillet, Bernardin, Portrat, Vergauwe, & Camos, 2007; Barrouillet & Camos, 2012; Barrouillet & Camos, 2015; Barrouillet et al., 2017; Brown, 1958; see also Ricker, Vergauwe, & Cowan, 2016; Vergauwe & Cowan, 2014; Vergauwe & Langerock, 2017). The interference account posits that forgetting in immediate memory is due to interference among the items in the memory set (e.g., Oberauer & Lewandowsky, 2014). For instance, it was shown that performance on a working memory test declines (i.e., interference increases) as a function of the similarity between WM items (Conrad & Hull, 1964; Oberauer, Farrell, Jarrold, Pasiecznik, & Greaves, 2012), their proximity in a list (Henson, 1999), and decreased spatial distance between items in a visual array (Rerko, Oberauer, & Lin, 2014). On the other hand, proponents of the temporal decay or decay-and-refresh hypothesis, initially put forward by Baddeley and Hitch (1974; see also Baddeley, 1986; Brown, 1958), claim that memory traces in WM are forgotten because their activation suffers from the passage of time (Barrouillet & Camos, 2015; Raye et al., 2007; Vergauwe & Cowan, 2014; Vergauwe & Cowan, 2015). Decay and interference as causes for forgetting are probably not mutually exclusive; rather, they might contribute in varying degrees to forgetting, depending on the given circumstances of the WM operation (Barrouillet et al., 2017).

Within the framework of the temporal decay account, the decline of activation of WM representations over time can be prevented or reversed by a process called refreshing. WM refreshing refers to one proposed maintenance mechanism that counteracts forgetting, thereby retaining information in WM (e.g., Barrouillet et al., 2004; Barrouillet et al., 2007; M. K. Johnson, Reeder, Raye, & Mitchell, 2002). It is suggested that refreshing is a process whereby items that are currently in WM are reactivated by placing them in the focus of attention (e.g., Raye et al., 2007). According to Vergauwe and Langerock (2017), refreshing is initiated and accomplished by actively “thinking of“ the items in WM (see also Souza et al., 2015). Importantly, it has been claimed that refreshing, through the help of controlled attention (Engle et al., 1999), operates serially, focusing on only one item at a single moment in time and cycling through the individual items in WM space (Barrouillet & Camos, 2012; Barrouillet & Camos, 2015; Vergauwe et al., 2014; Vergauwe & Cowan, 2014).

It should be acknowledged that refreshing is only one of several proposed maintenance mechanisms. In particular, it should be clearly distinguished from verbal or articulatory rehearsal, which is the maintenance process within the phonological loop of the multicomponent model of WM (Baddeley & Hitch, 1974; Baddeley, 1986; Baddeley, Gathercole, & Papagno, 1998; Baddeley, 2003b). Verbal rehearsal is often described as inner or subvocal speech (Baddeley, 2003a) and people engage in this activity, for example, when they repeat names or a phone number in their heads. There are two main differences between refreshing and verbal rehearsal (Camos & Barrouillet, 2014; Vergauwe & Langerock, 2017). Firstly, these two maintenance mechanisms diverge in the feasible scope of the to-be-maintained material. Refreshing is a domain-general process; it can work on verbal as well as visuo-spatial content. In contrast, verbal rehearsal is restricted to verbal or speech-related information and is, as such, domain-specific. Secondly, as already mentioned, refreshing relies on attention (e.g., Barrouillet et al., 2004; Raye et al., 2007), while it is assumed that verbal rehearsal requires attention only minimally (e.g., Camos, Lagner, & Barrouillet, 2009; Camos & Barrouillet, 2014; Naveh-Benjamin & Jonides, 1984). The differentiation between these two maintenance processes was demonstrated in a wealth of behavioral, developmental, and neuroimaging studies (viz. Camos et al., 2009; Cowan et al., 1998; Hudjetz & Oberauer, 2007; Raye et al., 2007; Vergauwe et al., 2014). Nevertheless, as Vergauwe and Langerock (2017) pointed out, although refreshing and verbal rehearsal are probably independent mechanisms, they are not mutually exclusive and both can contribute to the maintenance of information (see also Camos & Barrouillet, 2014, for a detailed discussion about maintenance processes).

As already mentioned, according to the decay-and refresh account (e.g., Barrouillet et al., 2004; Barrouillet et al., 2007; Barrouillet & Camos, 2012; Barrouillet & Camos, 2015; Barrouillet et al., 2017; Brown, 1958; see also Ricker et al., 2016; Vergauwe & Cowan, 2014; Vergauwe & Langerock, 2017), refreshing keeps the activation of WM items high (Barrouillet et al., 2004; Barrouillet et al., 2007; M. K. Johnson et al., 2002). Most research that investigated refreshing concentrated on the effects of refreshing on recall or recognition memory performance at the end of the trial or session. For example, several studies (e.g., Barrouillet et al., 2004; Barrouillet et al., 2007; Hudjetz & Oberauer, 2007) used variants of the complex span task, asking participants to memorize certain information (e.g., words, alphanumeric characters, locations, etc.) while performing processing tasks (e.g., counting backwards in steps of three or judging the parity of digits) between the presentation of the memory items. These types of complex span tasks, entailing both maintenance and processing components, are at the heart of the time-based resource-sharing model or TBRS (e.g., Barrouillet et al., 2004; Barrouillet & Camos, 2015), which claims that WM performance depends on the cognitive load (CL) of the complex span task (see also Vergauwe et al., 2014). CL is defined as the proportion of time between the presentation of the memoranda that is devoted to the distracting concurrent task. Because it is assumed that both maintenance and processing rely on the same resource of central attention that has to be shared in a time-based, sequential way (Barrouillet et al., 2004; Barrouillet et al., 2007; Camos et al., 2009; Vergauwe et al., 2014), the more time is devoted to the processing task, the less time is available for refreshing the memoranda. In line with that, as more time is devoted to refreshing (i.e., low CL), the memory span (i.e., memory performance) increases (Barrouillet et al., 2004; Barrouillet et al., 2007; Hudjetz & Oberauer, 2007; Vergauwe, Barrouillet, & Camos, 2010). Also, it was shown that memory performance increases as a function of the number of times an item was refreshed (Souza et al., 2015).

Although it is known that items in the focus of attention have a special status in terms of heightened accessibility (e.g., Garavan, 1998; Oberauer, 2002; Oberauer & Hein, 2012), it was only demonstrated recently that this also applies to attentional refreshing (Vergauwe & Langerock, 2017). A recent study by Vergauwe and Langerock (2017) examined the local effects of refreshing on WM representations. That is, rather than looking at the effects of refreshing on the final memory outcome, the effects of refreshing on reaction times (RT) were examined. In their paradigm (Vergauwe & Langerock, 2017), participants were presented with four letters within four boxes (also called memoranda), shown sequentially and each for either 1000 (Exp. 1, 3, and 5), 500 (Exp. 2), or 350 ms (Exp. 4). One box was presented in each quadrant of the display and the order of presentation was always as follows: top left, top right, lower left, lower right. Participants were asked to memorize these letters. After a briefly presented pattern mask, one to five refreshing cues in the form of a red highlighted box were shown in the instructed-refreshing condition, also always in the same order as the presentation of the memoranda and each for 1000 ms. For each refreshing cue, subjects were instructed to think of the memorandum that was previously presented at the position of the red box. Thus, it was under experimental control which items were refreshed. Lastly, a probe letter appeared and participants had to indicate as quickly as possible whether this letter corresponded to one of the four memoranda or not. Importantly, the probe letter could either match the just refreshed memorandum, one of the other three memoranda, or neither of the four memoranda. The important finding was that people's RT for probes that matched the just refreshed memorandum was shorter than the RT for probes that matched any of the three not refreshed memoranda. Vergauwe and Langerock (2017) concluded that this refresh benefit provides clear evidence that WM information in the focus of attention benefit from a privileged state in terms of a high accessibility, supporting the feasibility of the local effects to study refreshing. Given these convincing findings, we decided upon using these local effects, instead of global memory accuracies, as a measure of refreshing in our experiments.

In summary, refreshing is a recently proposed maintenance mechanism of WM that increases the activation of items in WM by bringing them into the focus of attention (e.g., Raye et al., 2007). Refreshing can be initiated by thinking about the WM representations (Souza et al., 2015; Vergauwe & Langerock, 2017). Importantly, refreshing increases the accessibility of the refreshed items, which can ultimately be observed in shorter RTs (Vergauwe & Langerock, 2017).

The Current Study

The central purpose of our investigation was to test whether the AB (i.e., WM consolidation) affects the ability to refresh. Specifically, is the refresh benefit during the AB, compared to after the AB, reduced or not? We reasoned that a dual-task paradigm, in which consolidation and refreshing had to be performed concurrently, would provide an appropriate way to examine this. With their instructed-refreshing paradigm, Vergauwe and Langerock

(2017) provide a means to control which WM representation is refreshed at which point in time. However, we needed to make several adjustments and simplifications to this paradigm, in order to make it viable to function as part of our dual-task paradigm. The feasibility of these modifications were tested in the first experiment.

The major differences of our simplified instructed-refreshing paradigm to the paradigm by Vergauwe and Langerock (2017) were that we presented the memoranda simultaneously instead of sequentially, in order to get rid of any advantages the last presented memorandum might have in terms of a higher accessibility and shorter RTs (see Nee & Jonides, 2008; Öztekin, Davachi, & McElree, 2010; Vergauwe et al., 2016). That is, since the activation of an item decreases as a function of the time that has passed since its presentation, we equalized the initial activations of the memoranda by presenting them simultaneously. Moreover, we only presented two instead of four memoranda and we included only a single refreshing cue in our paradigm. This modification was essential because it allowed a clear distinction between cuing within and outside of the AB in the dual-task paradigm. If we would have used multiple cues for multiple memoranda, the prolonged presentation time of cues would not have clearly fitted within or outside the temporal boundaries of the AB. The first experiment served the purpose of verifying that the modifications we made yield similar refresh effects as those found by Vergauwe and Langerock (2017) and, if that were the case, to inform us about the time interval between the refreshing cue and the probe that corresponds to the largest refresh benefit. This is essential because this would demonstrate the feasibility of our modified instructed-refreshing paradigm as one part of the dual-task paradigm in the second experiment. To anticipate, the results of the first experiment showed that our simplified instructed-refreshing paradigm is viable, replicating the refresh benefit observed by Vergauwe and Langerock (2017). Hence, we adopted a modified version of this simplified instructed-refreshing paradigm in the dual-task paradigm of the second experiment.

In the second experiment we examined to what extent refreshing is possible during consolidation (i.e., during the AB window). It was recently proposed that both consolidation and refreshing require controlled attention as a resource (Ricker et al., 2018). If that is true, we would expect that these two processes interfere with each other when performed concurrently. On the other hand, we would not expect interference or performance decrements if these two processes rely upon separate resources. In order to probe these two competing predictions, we created a dual-task paradigm (see Pashler, 1994, for a discussion), in which consolidation and refreshing had to be done concurrently. In more detail, two letters (memoranda) had to be memorized, one of which was later cued for refreshing. Critically, we manipulated whether participants had to refresh the memoranda during or after the AB window, elicited by the consolidation of a picture (T1). The time course of the AB due to the picture was determined in a separate block of trials. We hypothesized that the refresh benefit (i.e., the difference in RT between not just refreshed and just refreshed items) during the AB is reduced if consolidation and refreshing require the same resource. On the other hand, we hypothesized that the refresh benefit within the AB is not reduced if the two processes utilize different resources.

There might be a mechanistic account for the possibility that refreshing is unaffected by ongoing consolidation, at least in our design. As already mentioned, presenting a cue ahead of T2 in an AB paradigm has the effect that attention is reengaged so that T2 can be properly identified and reported (Nieuwenstein et al., 2005; Nieuwenstein, 2006; Nieuwenstein et al., 2009; Vul et al., 2008; Wyble et al., 2009; Wyble et al., 2011; Zhang et al., 2008). The delay of reengaging attention is approximately 100-200 ms (Nieuwenstein et al., 2005). If one conceives of the refreshing cue in our paradigm as equivalent to the cue in connection with AB paradigms, it becomes apparent that the time interval between the refreshing cue and the probe of 800 ms is much higher than the delay estimate of reengaging attention. As a consequence, independent of whether the refreshing cue is presented within or after the AB, there would probably be plenty of time for the refreshing cue to reengage attention for refreshing, provided that refreshing relies upon the same attentional mechanisms as those which are involved in the effects of precuing T2 in AB paradigms. As a result, there would probably no difference of refresh benefits inside and outside of the AB. Of course, one should bear in mind the important difference that cuing in the traditional AB paradigm refers to a newly externally presented stimulus, whereas the refreshing cue in our paradigm would pertain to an internal WM representation (M. R. Johnson et al., 2013).

Experiment 1

Method

Participants. Participants were 21 undergraduate psychology students (12 female, 9 male) from the University of Groningen. Their age ranged from 18 to 26 years (age in years: M = 20.3, SD = 2.2). As a first-year course requirement, participants completed the experiment in return for course credits. This experiment was approved by the ethical committee psychology (ECP) from the University of Groningen.

The decision of collecting data for 21 participants rested upon an a priori power analysis. We conducted a pilot experiment, with eleven participants, that was exactly the same as the first experiment, described here. We were interested in the interaction effect between Condition and Stimulus onset asynchrony (SOA) and the difference between the two conditions (i.e., Just Refreshed [JR] vs. Not Just Refreshed [NJR]) at the longest SOA of 800 ms. The effect size of the former effect was η_p^2 = 0.216, whereas the effect size of the latter was d_z = 0.752, which represents Cohen's d for difference scores in correlated designs, as described in Lakens (2013).[1] Based on these two effects the required sample sizes were determined. The power calculation for the interaction effect was based on α = .05, power = .95, and the Greenhouse-Geisser (Greenhouse & Geisser, 1959) non-sphericity correction ε =

.681; the required sample size was n = 17. The power analysis for the dependent-samples t-test was based on α = .05, power = .95, and a one-tailed alternative hypothesis (H_a: JR – NJR < 0) and calculations yielded a sample size of n = 21. Thus, choosing the largest of the two sample size estimates, we decided to collect data for 21 participants. The power analyses were performed using the software G*Power (Faul, Erdfelder, Lang, & Buchner, 2007).

Apparatus and stimuli. The stimuli in the experiment were fixation crosses, rectangles, lower- and uppercase letters, and masking stimuli. Both lower- and uppercase letters were presented in a non-bold Courier New font. Lowercase letters had a font size of 48 pt, whereas uppercase letters had a font size of 32 pt. We restricted the sample of letters to only contain consonants and we further excluded the letters ‘W’, ‘Y’, and ‘Z’, both for lower- and uppercase letters. All uppercase letters were surrounded by rectangles, measuring approximately 1.6 cm in width and 2.2 cm in height. The borders of these rectangles were colored with thin (2 pixels) black lines. In the absence of any letters, the borders of rectangles could also be colored with thick (6 pixels) red lines. The uppercase letters were followed by a masking pattern at the same position, consisting of three overlaid uppercase letters (i.e., A, I, and O; see Figure 1). This masking pattern was surrounded by black rectangles, as well. This kind of backward masking, where the masking pattern stimulates the same regions of the visual cortex as the letter, was used to interrupt re-entrant processing in visual cortex (Fahrenfort, Scholte, & Lamme, 2007; see also Breitmeyer & Öğmen, 2006 for a great overview). Importantly, lowercase letters were not surrounded by rectangles. Both lower- and uppercase letters and the masking pattern had a black color and were presented on a white background (see Figure 1).

The experiment was conducted in a darkened room. Instructions and stimuli were presented on a 19-in. CRT monitor with a resolution of 1024 × 768 pixels and a refresh rate of

100 Hz. The experiment was programmed and stimuli were presented in E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA).

Design and procedure. We employed a 3 (Condition: JR, NJR, and New) × 4 (SOA: 100, 200, 400, and 800 ms) full-factorial design. Both factors were within-subjects variables. Importantly, however, the probabilities of the levels of Condition were not balanced, with probabilities of 0.25 for JR, 0.25 for NJR, and 0.5 for New. The reason for this distribution was to avoid a response bias towards participants' decision that the probe was presented before (i.e., pressing the 'left arrow' key; see Figure 1). Since both JR and NJR corresponded to letters that were presented before, there would be an imbalance of pressing the 'left arrow' and 'right arrow' keys if equal proportions were assigned to each of the three conditions, as done by Vergauwe and Langerock (2017, Exp. 1-5). Instead, with our design, there was an equal probability that the probe was presented before or not presented before (see below).

At the beginning of a test session, participants had to read and sign the informed consent form. After that, they were seated at a comfortable distance in front of the computer in a private cabin. Short and standardized verbal instructions were provided by the researcher. Subsequently, participants were allowed to start the experiment. Detailed instructions were shown on the screen, where it was explicitly stated several times that subjects could contact the researcher at any time, in case of questions. Once participants read the instructions, we tested their comprehension of the instructions. This was done by sequentially presenting schematic pictures of all six possible trial outlines for which a response, indicating the hypothetical correct answer, had to be made. This comprehension test did not require speeded responses and accuracy feedback was provided for each trial outline. Following that, participants had to complete 32 practice trials. At the end of each practice trial, participants received feedback about their performance. However, only information about the accuracy but not the reaction time (RT) was provided. Collapsed over all SOAs, participants had to reach an accuracy higher than 0.65 in every condition in order to be allowed to proceed to the experimental trials. Otherwise, the researcher verbally explained the task once again with the aid of a visual depiction of the six possible trial outlines, and participants had to read the instructions and perform the comprehension test and the practice trials anew. In total, participants had to do 320 experimental trials. After the experimental trials, participants were fully debriefed, both verbally and in written form.

The full trial sequence can be inspected in Figure 1. At the start of a trial, participants saw a fixation cross at the center of the screen. Participants had to press the spacebar to initiate the trial, after which the fixation cross remained on the screen for an additional duration of 500 ms. Then, two boxes appeared on the screen; one on the right and one on the left of the fixation cross. Within these boxes were two randomly chosen uppercase letters, with the restriction that they were not the same. Together with the central fixation cross, this display remained for 2000 ms. The boxes with the letters inside were positioned quite close to the fixation cross, so that they were near the central visual field. In that way, participants could fixate their gaze on the fixation cross and still identify the adjacent letters (e.g., Levi, McGraw, & Klein, 2000). Participants were asked to memorize the letters. Subsequently, the uppercase letters were replaced by a masking pattern, as described above, while the boxes and the fixation cross remained unchanged. This masking display was shown for 50 ms. In the next display, the mask was removed from the boxes. However, either the left or the right box, with equal probability and balanced over all design cells, was highlighted with thick red borders and subjects had to think about the letter that was just presented at the position of this highlighted box. This configuration was shown for 100 ms, which was then replaced by a blank screen that was shown for either 0, 100, 300, or 700 ms, resulting in SOAs of 100, 200, 400, and 800 ms, respectively; each duration occurred equally often but was selected at random for each trial. Lastly, a lowercase letter, serving as the probe, was presented at the center of the screen. This letter belonged to one of three categories: In 25% of the trials, this letter was the same as the letter presented at the position of the red box; this constituted the JR condition. In 25% of the trials, the letter was the same as the letter presented at the position of the non-red box; this was NJR condition. In 50% of the trials, the letter did not correspond to either of the two letters within the two boxes, thus, belonging to the 'new' condition. The 'left arrow' key had to be pressed if the probe corresponded to one of the two letters, while the 'right arrow' key had to be pressed if it did not. Participants were instructed to respond as quickly and accurately as possible. Importantly, the probe display was terminated when a response occurred, with a maximum response window of 2000 ms. If no response was made within this time window, the response was considered an error.

Data processing and statistical analyses. We excluded participants that had an accuracy lower than 0.5 (chance level) in any of the cells of our design. One participant failed to reach sufficient accuracy and was, thus, not considered in further analyses.[2] Specifically, for this participant accuracies in the New condition ranged from 0.05 to 0.35, which is very low compared to the other participants. The average accuracy for the reduced sample of 20 participants over all trials was 0.94 (*SD* = 0.04). Moreover, we excluded trials for which an incorrect response was made (also trials for which no response at all was made).

For the remaining correct trials, we performed an outlier elimination procedure (cf. Jolicœur & Dell'Acqua, 1998; Van Selst & Jolicœur, 1994). RT distributions have the robust tendency to be skewed to the right (Heathcote, Popiel, & Mewhort, 1991; Ratcliff, 1993; Ulrich & Miller, 1994). Extremely short or long RTs might not solely (or not at all) represent the cognitive processes of interest but also non-relevant factors; for example, lapses of thought, eye blinks, and inattentiveness (Heathcote et al., 1991; Ratcliff, 1993). The exclusion of such trials is not straightforward. As Heathcote and colleagues (1991) argued, there are two explanations for skewed RT distributions: Either the process of interest yields skewed distributions, or skewness is due to nuisance variables solely or on top of the process of interest. Hence, we can never be certain whether any exclusion procedure eliminates the right trials. However, in order to exclude trials where such non-relevant factors were probably involved as good as possible, we performed the modified recursive outlier elimination procedure, advocated by Van Selst and Jolicœur (1994; see also Jolicœur & Dell'Acqua, 1998), for each cell of our design. In essence, this procedure excludes observations iteratively. For each iteration there is a criterion cut-off (C), which depends on the sample size: the smaller the sample size, the higher C (cf. Van Selst & Jolicœur, 1994). C-values for some specific sample sizes were provided by Van Selst and Jolicœur (1994); we calculated C-values for missing sample sizes through linear interpolation. The steps of this procedure were as follows: First, the maximum RT was removed from the sample. Based on this sample, the mean and standard deviation were calculated. The low and high thresholds were obtained by calculating the difference and the sum of the mean and the product of C and the standard deviation (Jolicœur & Dell'Acqua, 1998). If the minimum and maximum values from the original sample exceeded these thresholds, the corresponding value was removed from the sample, which then constituted the new sample for the next iteration. This procedure was repeated until no values exceeded the thresholds anymore or until the sample size was smaller than five. Based on the resulting data, we calculated aggregate (mean) RTs for each subject and each cell of our design, which were used for any further statistical analyses.

We performed a repeated measures analysis of variance (ANOVA), with type III sum of squares, and four pooled dependent-samples *t*-tests, with RT as the dependent variable. Means and heterogenous[3] between-subjects standard errors of the corresponding group means can be inspected in Figure 2. For all model terms of the within-subjects ANOVA we calculated the generalized eta squared (η_G^2) as the effect size (Bakeman, 2005; Olejnik & Algina, 2003). It has been argued that the generalized eta square effect size is superior to the more commonly reported partial eta square effect size because it allows better comparisons between uncorrelated and correlated group designs (Bakeman, 2005). Also, for all pooled dependent-samples *t*-tests, we report an adjusted Cohen's *d* effect size, as described in Dunlap, Cortina, Vaslow, and Burke (1996). In case Mauchly's test indicated a violation of sphericity (α = .05) for any model term in the repeated measures ANOVA, we used the conservative Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) instead of the liberal Huynh-Feldt correction (Huynh & Feldt, 1976) for that specific term. For all statistical tests we employed an alpha level of α = .05. The outlier exclusion procedure was executed in R (R Core Team, 2018) and statistical analyses were performed in R (R Core Team, 2018), using the "afex" package (Singmann, Bolker, Westfall, & Aust, 2018).

Mauchly's test of sphericity did not indicate a violation of sphericity for the main effect of SOA and the interaction between SOA and Condition (*W*[5] = 0.61, *p* = .124; *W*[5] = 0.65, *p* = .171, respectively). However, Shapiro-Wilk tests (Shapiro & Wilk, 1965) indicated violations of normality for JR at an SOA of 400 ms (*W*[20] = 0.87, *p* = .014) and JR at an SOA of 800 ms (*W*[20] = 0.90, *p* = .039). The remaining six tests were non-significant (all *p*s > .200). Because the Shapiro-Wilk test (Shapiro & Wilk, 1965) was criticized for the characteristic of having an inflated type 1 error rate, we also examined histograms and quantile-quantile (qq) plots. The RT distributions within the eight design cells looked reasonably normal. In addition, it has been argued (although not uncontroversial) that an ANOVA is robust against small violations of normality (Glass, Peckham, & Sanders, 1972; Lunney, 1970). Hence, we did not apply any data transformations or non-parametric tests.

Results

The main effect of Condition was significant, $F(1, 19) = 23.57$, $p < .001$, $\eta_G^2 = 0.054$. Also, the main effect of SOA was significant, $F(3, 57) = 29.93$, $p < .001$, $\eta_G^2 = 0.060$. As can be seen in Figure 2, RTs tended to decrease when SOA increased. Against our expectation, however, the data did not indicate unequal differences between JR and NJR at the various SOAs, which was evident in the lack of an interaction between Condition and SOA, $F(3, 57) = 0.48$, $p = .697$, $\eta_G^2 = 0.001$. Even though the interaction turned out not to be significant, we performed one-sided pooled dependent-samples *t*-tests, comparing RTs for both conditions at each SOA. Differences between conditions were significant at each SOA (100 ms: $t[19] = -3.50$, $p = .001$, $d = 0.348$; 200 ms: $t[19] = -4.10$, $p < .001$, $d = 0.473$; 400 ms: $t[19] = -3.98$, $p < .001$, $d = 0.511$; 800 ms: $t[19] = -3.06$, $p = .003$, $d = 0.500$), with shorter RTs for JR at each SOA.

Discussion

We replicated the findings by Vergauwe and Langerock (2017) of a local refresh benefit, evident in RTs, thereby demonstrating the feasibility of our simplified instructed-refreshing paradigm. Surprisingly, we found clear signs of refreshing at all SOAs, which were evident in differences in RTs between the JR and NJR condition (cf. Vergauwe & Langerock, 2017). Modern literature on refreshing tends to employ long durations for refreshing, for example 500 ms (Souza et al., 2015) or 1000 ms (Vergauwe & Langerock, 2017). Our results hint towards the possibility that less time for refreshing would be equally effective in eliciting refreshing effects. Moreover, although absolute RTs decreased as SOA increased (see Figure 2), we did not find unequal differences in RTs at the four SOAs. This implied that any of these SOAs could be used in the dual-task paradigm of the second experiment. However, since the pilot study indicated the presence of an interaction, with increasing refresh benefits as SOA increased, we decided upon avoiding any risks and to use an SOA of 800 ms in the second experiment.

Experiment 2

Methods

Participants. We sampled 33 participants (27 female, 6 male). Their age ranged from 18 to 25 years (age in years: *M* = 20.8, *SD* = 2.2). Of these, 20 (17 female, 3 male) were undergraduate psychology students at the University of Groningen, taking part in the experiment as a first-year course requirement. Their age ranged from 18 to 22 years (age in years: *M* = 19.6, *SD* = 1.4). The remaining 13 subjects (10 female, 3 male) were paid 8 € for participation. Their age ranged from 20 to 25 years (age in years: *M* = 22.7, *SD* = 1.8). The study was advertised at the website for paid studies of the University of Groningen. Anyone registered in this system could enroll for the study. This experiment was approved by the ethical committee psychology (ECP) from the University of Groningen.

Our sample size rested upon a recent simulation study, which demonstrated that on average 30 participants would be needed to obtain a Bayes factor smaller than 0.333 if no effect is actually present (Schönbrodt, Wagenmakers, Zehetleitner, & Perugini, 2017). At first, we exhausted the possibility to sample from the first-year psychology participant pool. Once we gathered the data for these 20 participants, we analyzed the data and decided upon testing more participants. Subsequently, we made use of the paid participant pool to collect additional data for 13 participants. One reason we used Bayesian statistics is the possibility to collect more data even after data inspection (Schönbrodt et al., 2017). The frequentist statistical framework would not allow for such a step-wise data collection and analysis (see Wagenmakers et al., 2018, for a detailed discussion on advantages of Bayesian inference).

Apparatus and stimuli. The experiment consisted of two blocks of trials, which we call the AB block and the combination (AB and refreshing) block. In both blocks we used fixation crosses, masking stimuli, pictures as T1, and uppercase letters embedded within black rectangles as memoranda. Furthermore, in the AB block we presented digits as T2 and verbal prompts, asking participants which memoranda they have seen. In the combination block, red rectangle served as refreshing cues (T2) and lowercase letters as memoranda probes. The lower- and uppercase letters, as well as the black and red rectangles and the masking pattern for the uppercase letters were the same as described in the first experiment. In addition to the black fixation crosses described in the first experiment, we included a bright blue one. We derived 200 pictures from the database by Brady, Konkle, Alvarez, and Oliva (2008). More specifically, we used the state pair pictures, which showed 100 simple objects in two slightly different states, yielding 100 pairs of pictures. All pictures had 256 × 256 pixels with an approximate display size of 2.7 × 2.7 cm. As the lowercase letters in the first experiment, the digits (0-9) were shown in a 48 pt Courier New font. The mask for the digit consisted of four overlaid digits (i.e., 0, 2, 4, and 5), also in a 48 pt Courier New font. The verbal prompts (“Left letter?” and “Right letter?”) were presented in an 18 pt Courier New font.

The execution of the experiment was done in a dimly lit cabin. The construction of the experiment and the presentation of stimuli were done using the E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). Instructions and stimuli were shown on a 19-in. CRT monitor with a resolution of 1028 × 768 pixels and a refresh rate of 100 Hz.

Design and procedure. The experiment consisted of two blocks of trials. Although all participants had to complete both the AB and the combination blocks, their order was counterbalanced across participants. For the AB block, we employed a one-way (SOA: 300, 700, 1100, 1500, and 1900 ms) within-subjects design. The combination block consisted of a 3 (Condition: JR, NJR, and New) × 2 (Picture-cue onset asynchrony [PCOA]: 300 and 1200 ms) full factorial design, with both factors being within-subjects variables. Here, as in the first experiment, the probability of occurrence for JR and NJR was 0.25 each and 0.5 for New (cf. Vergauwe & Langerock, 2017, Exp. 1-5), for the purpose of avoiding a response bias. A description of these conditions and their probability distribution can be found in the first experiment.

The procedure for the informed consent and the provision of general instructions was the same as in the first experiment. Detailed instructions about the specific blocks were presented directly before the execution of each block, instead of presenting all information at once at the beginning of the experiment. In the AB block, participants first had to complete ten practice trials. No feedback was provided and no minimum accuracy restriction was applied. Subsequently, participants had to complete a total of 100 experimental trials. The combination block started with a comprehension check of the block-specific instructions. Similar to the first experiment, visual depictions of all six trial outlines were shown sequentially on the screen and participants had to indicate the correct response. For each visual depiction feedback was provided about participants' accuracy. No speeded responses were required. Subsequently, 32 practice trials had to be completed. In order to be allowed to proceed to the experimental trials, subjects had to reach an accuracy higher than 0.6 for each condition, collapsed over all PCOAs. If they were below this criterion, the researcher verbally, and with the aid of a visual depiction of the trial outlines, explained the tasks again. Then, the whole practice procedure, including reading the instructions, had to be done again. In total, 160 experimental trials had to be performed. As an exception, the first three participants completed 192 experimental trials. However, we noticed that the experiment took too long and, hence, reduced the number of experimental trials to 160 for the remaining participants. After completion of both blocks, participants were fully debriefed in verbal and written form.

The full trial sequence of the AB block can be inspected in Figure 3. The start of the trial, from the presentation of the fixation cross until, and including, the memoranda mask, was the same as in the first experiment. In line with that, instructions to memorize the memoranda were the same as before. The memoranda mask was followed by a screen where only the black boxes were shown. This display remained for 700 ms. In general, we decided to include the empty boxes in the blank displays to prevent any unnecessary visual disturbances due to repeated on- and offsets of the boxes that might affect task performance. Subsequently, a randomly chosen picture was presented between the two boxes at the center of the screen for 50 ms. Participants had to memorize this picture in as much detail as possible. This picture was the first target (T1) in this adapted AB paradigm. After that, another screen, containing only the boxes, was shown for 250, 650, 1050, 1450, or 1850 ms, yielding SOAs of 300, 700, 1100, 1500, and 1900 ms, respectively. In the next display, a randomly chosen digit, serving as T2, was inserted between the two boxes at the center of the screen. Participants had to identify this digit, which was shown for 60 ms, in order to be able to report it at the end of the trial. The digit backward mask was presented at the same position as the digit for 400 ms. It was demonstrated that a mask for T2 is necessary to observe an AB (Giesbrecht & Di Lollo, 1998; Vogel & Luck, 2002) However, it is worth mentioning that this sort of backward mask probably did not fully prevent consolidation; in contrast, it was demonstrated that consolidation continues long after the presentation of a backward mask (Nieuwenstein & Wyble, 2014). Next, the texts “Left letter?” and “Right letter?” were presented in that order at the center of the screen, prompting participants to type in the left and right memoranda, respectively. If they were not able to remember the memoranda, they could press ‘enter’. Thereafter, the picture recognition test was implemented by presenting the two state-pair pictures (Brady et al., 2008) next to each other, one on the left and one on the right. One of these pictures was the picture presented before as T1, while the other one was a foil. The position (left vs. right) of old and foil pictures was balanced across design cells; that is, old pictures occurred equally often on the left and on the right. However, within this restriction, the order of location was selected randomly for a given trial. Subjects had to indicate which picture they saw earlier by pressing either the ‘left arrow’ or ‘right arrow’ key. Lastly, a bright blue cross was presented at the center of the screen. At that point, participants were asked to type in the digit (T2) they saw earlier. The ‘enter’ key could be pressed in case participants were not able to recall the digit. For all these tasks (i.e., typing in the left and right memoranda, recognizing the picture, and recalling the digit) no time pressure was put upon participants.

The structure of trials in the combination block can be seen in Figure 4. The start of the trials in the combination block was identical to the AB block; until, and including, the picture (T1) display everything was exactly the same. The blank screen, containing only the boxes, that followed the display of the picture had a duration of either 250 or 1150 ms, resulting in PCOAs of 300 and 1200 ms, respectively. Subsequently, one of the two boxes was highlighted with thick red borders. The selection of the left or right box for highlighting was balanced across the design cells and, within this restriction, the order of location was random for a given trial. This display remained on the screen for 100 ms. Participants were asked to think about the letter (memorandum) that was just presented at the position of the red box. Thus, the red box served as the refreshing cue (cf. Vergauwe & Langerock, 2017), which was replaced by a blank screen with the usual black boxes for 700 ms. Identical to the first experiment, a lowercase letter, serving as the probe, was then presented at the center of the screen. This probe either belonged to the JR, NJR, or New condition. Participants had to indicate whether the probe was one of the memoranda they had seen earlier. To this end, they had to press the ‘left arrow’ key if they had seen the letter earlier and the ‘right arrow’ key if they had not. Subjects had to respond as quickly and accurately as possible, with a maximum response window of 2000 ms. The program proceeded to the next display once a response has been made or 2000 ms have passed. If participants failed to respond, this was counted as an error.

Data processing and statistical analyses. We excluded three participants from the data of both blocks.[4] One participant had a mean accuracy of zero for the recall of the left and right memoranda and the identification of digits (T2) in the AB block. Furthermore, this participant had an extremely low mean accuracy of 0.31 for reporting the probe in the combination block, compared to the other participants. Lastly, for the picture (T1) recognition in both blocks, this participant always responded with the 'left arrow' key. This pattern of responding with the 'left arrow' key for the picture recognition was also found for the other two participants. The response styles of these three participants suggest that the task was not taken seriously or was at least deeply misunderstood. After excluding these three subjects, we arrived at a sample size of 30. In the combination block, the remaining 30 participants reached an overall mean accuracy of 0.95 (SD = 0.04) for reporting the probe.

For the AB block, we excluded trials in which the picture (T1) was not correctly recognized, whereas trials in which the probe letter was not correctly identified (including no response at all) were removed from the data for the combination block. For the remaining trials in the combination block we performed the outlier elimination procedure, as described in the first experiment (see Jolicœur & Dell'Acqua, 1998; Van Selst & Jolicœur, 1994). This outlier elimination procedure was unnecessary and, thus, not applied for the AB block since RTs were not of interest. After excluding the outliers from the combination block, we further abolished trials in which the picture (T1) was not correctly recognized. Based on the resulting data, we calculated aggregate (mean) values for each participant and each cell of the designs for the AB and combination blocks. These aggregates were used for our statistical analyses.

In contrast to the first experiment, we used Bayesian statistics to analyze data for the two blocks. Since we wanted to evaluate the evidence for both the null and alternative models, we calculated Bayes factors (BF), which essentially are a ratio of the likelihood of the data given the alternative model and the likelihood of the data given the null model, or, more formally, the ratio of the posterior and prior odds (e.g., Kass & Raftery, 1995; Rouder, Morey, Speckman, & Province, 2012; Wagenmakers, 2007). In particular, we performed Bayesian repeated measures ANOVAs (Rouder et al., 2012) on digit identification accuracy in the AB block, probe RT in the combination block, and picture recognition accuracy in both blocks.

Furthermore, we calculated inclusion Bayes factors (BF_{incl}) across matched models, similar to the procedure described by Mathôt (2017). To be precise, our BF_{incl} contrast models that include the effect of interest with models that do not contain the effect of interest. Importantly, higher-order interactions are exempted from this comparison. Mathematically, the BF_{incl} for a specific effect is then the sum of the normal BFs for models in which the effect of interest is present divided by the normal BFs for the same models, but this time not including the effect of interest.[5] Again, models that include higher-order interactions are not considered. The interpretation of our BF_{incl} is different from the interpretation of the normal BF, with the former referring to a specific effect and the latter to a model. In essence, the BF_{incl} allows us to judge the importance of a specific effect.

Moreover, Bayesian dependent-samples one- and two-sided *t*-tests (Rouder, Speckman, Sun, Morey, & Iverson, 2009) were calculated on the same dependent variables. To avoid any confusion, it is worth mentioning that Bayesian dependent-samples one-sided *t*-tests can have two different null models: For one, the null model can be that the difference between two groups or measurement occasions is zero. Alternatively, the null model could be the opposite of the alternative model; in that case, if we hypothesize that the difference is larger than zero, the null model would be that the difference is smaller than zero, and vice versa. This last option more closely resembles the familiar dependent-samples one-sided *t*-test in frequentist statistics. However, unless otherwise specified, we decided to use the former null model when doing a Bayesian dependent-samples one-sided *t*-test, stating that the difference between the two measures is zero.[6] All Bayesian dependent-samples *t*-tests are accompanied with the Cohen's d effect size (cf. Dunlap et al., 1996), as described in the first experiment. Means and the corresponding heterogenous between-subjects standard errors of the means for digit identification accuracy in the AB block and probe RT in the combination block can be inspected in Figure 5. Data processing and statistical modelling were done in R (R Core Team, 2018) and the "BayesFactor" package (Morey & Rouder, 2015). For both the Bayesian repeated measures ANOVAs and the dependent-samples one-sided *t*-tests the default settings were used (Morey & Rouder, 2015).

Results

AB block. We found decisive evidence for the model including SOA on digit (T2) identification accuracy (BF = 1.251×10^{54}).[7] As long as the AB is present, T2 identification accuracy should increase as SOA increases. Given this unidirectional prediction, we performed Bayesian dependent-samples one-sided *t*-tests of digit identification accuracy between successive SOAs. We found large performance differences up until 1100 ms, after which the performance reached a plateau ($BF_{700 – 300} = 3.427 \times 10^{14}$, $d = 2.531$; $BF_{1100 – 700} = 3.424 \times 10^{6}$, $d = 1.343$; $BF_{1500 – 1100} = 0.986$, $d = 0.295$; $BF_{1900 – 1500} = 0.557$, $d = 0.197$, respectively). Thus, the duration of the AB was at least 1100 ms. This can be further inspected in Figure 5, showing that pictures elicited an enormous AB, both in magnitude and duration. The range of accuracies across the different SOAs is very large, with tremendous identification deficits at the short SOAs. Looking at picture (T1) recognition accuracy, we found anecdotal evidence against the model including SOA (BF = 0.785; see Figure 6).

Combination block. Concerning probe RT in the combination block, we obtained anecdotal evidence for the individual main effect models of Condition (BF = 1.128) and PCOA (BF = 1.665) and the combined main effect model, including Condition and PCOA (BF = 2.134). However, we found anecdotal evidence against the combined main effect and interaction model (BF = 0.774).[8] All four Bayes factors suggest that we did not find strong evidence for either the null model or the four alternative models since all Bayes factors were very close to one (see Figure 5). This is also reflected in the inclusion Bayes factors for Condition, PCOA, and the interaction of Condition and PCOA (BF_{incl} = 1.224, BF_{incl} = 1.785, BF_{incl} = 0.522, respectively), which are all very close to one, suggesting that including any one of these effects did not dramatically change the odds for or against the models including the effect, compared to the models not including the effect. Comparing JR and NJR at a PCOA of 300 ms, we only found anecdotal evidence for the hypothesis that RTs are lower for JR (BF = 1.958, *d* = 0.205), and we even found anecdotal evidence against that hypothesis at a PCOA of 1200 ms (BF = 0.411, *d* = -0.080).

For the Bayesian repeated measures ANOVA on picture (T1) recognition accuracy (see Figure 7), we obtained anecdotal evidence against the main effect model of Condition (BF = 0.792). Furthermore, we found moderate evidence against the main effect model of PCOA (BF = 0.310) and the combined main effects model (BF = 0.273). And we even found strong evidence against the combined main effects and interaction model (BF = 0.063). The inclusion Bayes factors are in line with that, suggesting that the data were 1.231 times less likely for models including the effect of Condition (BF_{incl} = 0.813), 3.069 times less likely for models including the effect of PCOA (BF_{incl} = 0.326), and even 9.472 times less likely for models including the interaction between Condition and PCOA (BF_{incl} = 0.106), compared to models not including the corresponding effect.

Discussion

The results for digit (T2) identification accuracy in the AB block indicate that the picture (T1) elicited a very strong AB, evident in both its magnitude and duration. That is, when the picture and the digit were separated by a short SOA, the digit did not have the capability to reengage attention in time because the ongoing consolidation of the picture suppressed any deployment of attention, so that the digit could not be consolidated into WM. On the other hand, as SOA increased it was more likely that the consolidation of the picture was already advanced, so that the digit was more and more capable of reengaging attention again. This finding is in accordance with other studies using pictures as T1 (e.g., Linde et al., 2016; Most et al., 2005; Most et al., 2007; Nieuwenstein & Kromm, 2017), with a prolonged AB probably being due to a higher information load of pictures compared to simple alphanumeric characters (Linde et al., 2016; see also Nieuwenstein & Wyble, 2014). As a result of the increased information load, pictures might have a higher processing demand and need longer to be encoded and consolidated into WM. Alternatively, this strong AB might result from the fact that participants had to remember details of the pictures, rather than simply the gist (see Nieuwenstein & Kromm, 2017).

What is most essential is that T2 identification accuracy was markedly different at SOAs of 300 (M = 0.20, SE = 0.03) and 1100 ms (M = 0.83, SE = 0.03). This is important because we used similar SOAs (to be precise, PCOAs of 300 and 1200 ms) between the picture (T1) and the refreshing cue (T2) in the combination block. Given this large difference in T2 performance between the two SOAs and the fact the we used the same pictures as T1 with the same instructions in the AB and combination blocks, we can assume that the PCOA manipulation in the combination block was successful. More specifically, we can be quite certain that the refreshing cue was presented within the AB window (i.e., during consolidation) when we used the short PCOA of 300 ms and outside of the AB window (i.e., after consolidation) when we used the long PCOA of 1200 ms. We also ensured that the two blocks were as similar as possible in terms of task complexity. That is, we also integrated memoranda in the AB block, which participants had to recall at the end of the trial.

Unfortunately, we could not properly test the predictions of the shared and separate resources accounts on probe RTs in the combination block due to the lacking refresh benefit at 1200 ms, which did not warrant a comparison of the refresh benefit at 300 ms to the refresh benefit at 1200 ms (see Figure 5). In particular, the refresh benefit at 300 ms should be reduced, compared to 1200 ms, if consolidation and refreshing rely upon the same resources. If, however, these two processes utilize separate resources, the refresh benefit at 300 ms should be equally strong as the refresh benefit at 1200 ms. Independent of which hypothesis is correct, refreshing should be possible at 1200 ms. If the shared-resource hypothesis is true, refreshing should be possible because we saw in the AB block that the AB was completed at 1200 ms and the shared resource should, thus, be available again. If the separate-resource hypothesis is true, refreshing should be possible a fortiori because the resource for refreshing should be callable independent of the AB. In short, due to the lacking refresh benefit at 1200 ms, we were not able to properly test our two competing hypotheses. This puzzling finding leaves open the question of why participants were unable to refresh at 1200 ms. We do not have a definitive answer to that question but some possible reasons of why that might have happened are discussed in the General Discussion.

Exploratory Analyses

In order to make more sense of the results of the second experiment, we conducted some follow-up exploratory analyses. The premise for the testability of shared and separate resources hypotheses was that we at least observe a refresh benefit at a PCOA of 1200 ms. Unfortunately, the evidence indicated that there probably was no difference between NJR and JR trials at 1200 ms, leaving us unable to test the two hypotheses. Indeed, when we compared the refresh benefit in the first experiment at an SOA of 800 ms with the refresh benefit in the second experiment at a PCOA of 1200 ms, we found moderate evidence that the refresh benefit in the first experiment was larger than in the second experiment (BF = 5.979, d = 0.705).

The possibility exists that the expected refresh benefit at 1200 ms only occurred under specific circumstances. For instance, it would be interesting to see if the RT patterns, and the magnitude of the refresh benefit at 1200 ms in particular, in the second experiment were different for trials in which the left memorandum was probed compared to trials where the right memorandum was probed. To this end, we conducted a 2 (Condition: JR vs. NJR) × 2 (PCOA: 300 vs. 1200 ms) × 2 (Location: left vs. right) Bayesian repeated measures ANOVA on RT. RT means and heterogenous between-subjects standard errors of means can be examined in Figure 8. It is readily obvious that there was a large difference in overall RT between the left and right probed letters; we obtained conclusive evidence for the main effect model of Location (BF = 1.677×10^{14}), also reflected in the inclusion Bayes factor for the effect of Location ($BF_{incl} = 3.796 \times 10^{14}$), with much shorter RTs for the left probed letter. In fact, a similar spatial bias towards the left was observed in several studies, investigating different phenomena (e.g., Bowers & Heilman, 1980; Dickinson & Intraub, 2009; Foulsham, Gray, Nasiopoulos, & Kingstone, 2013; Williams & Reingold, 2001; Zelinsky, 1996). Crucially, we did not find compelling evidence in favor for the inclusions of the two-way interactions of Condition and Location of probed letter ($BF_{incl} = 0.297$) and PCOA and Location of probed letter ($BF_{incl} = 1.194$) and the three-way interaction of Condition, PCOA, and Location of probed letter ($BF_{incl} = 0.655$), demonstrating that Location did not interact with any other effects. In summary, although we see a large difference in overall RTs between the left and right probed letters, Location most likely did not interact with PCOA and Condition in predicting probe RTs. Most importantly, for the long PCOA of 1200 ms, we did not find clear evidence that refreshing was possible, neither for the left (BF = 0.127, d = 0.075) nor the right (BF = 1.316, d = 0.202) probed letter.

Furthermore, we also tested whether a similar bias towards the left memorandum could be observed in the first experiment. Hence, we conducted a 2 (Condition: JR vs. NJR) ×

4 (SOA: 100, 200, 400, and 800 ms) × 2 (Location of probed letter: left vs. right) Bayesian repeated measures ANOVA on RT. Means and heterogenous between-subjects standard errors of means are presented in Figure 9. Here again, we found conclusive evidence for the main effect model of Location (BF = 1.265×10^{10}), which es corroborated by the decisive inclusion Bayes factor ($BF_{incl} = 2.324 \times 10^{15}$) for this effect.

Although we were not able to properly test the two competing hypotheses of whether consolidation and refreshing use the same or separate resources because of the lack of a refresh benefit at a PCOA of 1200 ms, we might be able to inquire these hypotheses by only considering a certain subgroup of participants. In particular, what if we only examine those participants who exhibited a refresh benefit at 1200 ms? We reasoned that it might have been possible that the task in the combination block of the second experiment was simply too complex for most participants, which is why we would observe a refresh benefit at 1200 ms only in a few participants. In line with that, it is well known that people differ markedly in their WM capacity and WM functioning in general (e.g., Bayliss, Jarrold, Gunn, & Baddeley, 2003; Daneman & Carpenter, 1980; Engle et al., 1999). To this end, we grouped participants into tertiles in terms of their difference in RT between NJR and JR at a PCOA of 1200 ms and only considered participants in the highest tertile (i.e., participants with the largest refresh benefit at 1200 ms). For this subgroup we performed the same 2 (Condition: JR vs. NJR) × 2 (PCOA: 300 vs. 1200 ms) Bayesian repeated measures ANOVA as already done for the complete set of participants. Naturally, because we only analyzed participants who did show a refresh benefit at 1200 ms, we found very strong evidence in favor of an effect of Condition ($BF_{incl} = 72.083$), indicating that RTs for JR and NJR differed markedly for these participants. Anecdotal evidence against the effect of PCOA was obtained ($BF_{incl} = 0.445$), showing that RTs were similar across PCOAs. Importantly, we found anecdotal evidence against the interaction between Condition and PCOA ($BF_{incl} = 0.711$). Even though we found a large refresh benefit at 1200 ms (BF = 265.599, d = 0.52), which in theory allowed us to probe the two hypotheses, the evidence against the interaction effect was negligible, prohibiting us to draw any firm conclusions in favor of either hypothesis.

To further explore the relationship between WM capacity and refreshing, we calculated an admittedly spontaneous and probably not optimal measure of WM capacity based on the data we had at hand. More specifically, we determined an overall performance score for each participant, that was based on performance on the recall of the left and right memoranda, picture recognition sensitivity, and digit identification accuracy in the AB block and picture recognition sensitivity in the combination block. We standardized the scores on these four performance measures and added the resulting z-scores. Subsequently, we standardized the resulting sum of z-scores once again. With this procedure, we hoped to obtain a relatively reliable and valid index of WM capacity. A Bayesian simple linear regression of WM capacity on the difference between NJR and JR at a PCOA of 1200 ms in the combination block was conducted. However, the evidence was in favor of the null model that there is no relationship between our index of WM capacity and the refresh difference (BF = 0.375).

General Discussion

The goals of the first experiment were to demonstrate the viability of our simplified version of the instructed-refreshing paradigm by Vergauwe and Langerock (2017) and to determine the optimal SOA between the presentation of the refreshing cue and the probe, both of which were essential for a successful implementation of this paradigm as part of the dual-task paradigm of the second experiment. We presented two memoranda simultaneously instead of four memoranda sequentially and varied the SOA between a single refreshing cue (cf. Vergauwe & Langerock, 2017) and the probe (see Figure 1). To this end, we compared the RTs for probes that either matched the JR or not NJR memorandum at four SOAs (100, 200, 400, and 800 ms) between the refreshing cue and the probe. In this way, we were able to control how much time was available for refreshing. Similar to the study by Vergauwe and Langerock (2017), we found reliable refresh benefits, indicating that our simplification of the instructed-refreshing paradigm is feasible. Astonishingly, at all SOAs participants were able to refresh the cued item, which was evident in large and consistent refresh benefits (see Figure 2), having the consequence that the specific choice of SOA in the second experiment would not matter that much.

The fact that we observed a refresh benefit at 100 ms indicates that people can refresh quite quickly. However, caution should be exercised to not interpret that finding of conclusive evidence that people need 100 ms or less to refresh. In theory, participants could have utilized not only the time between the refreshing cue and the probe for refreshing but also some unknown part of the time between the onset of the probe and the actual button press for refreshing. Although probably a large proportion of the time until the button press was devoted to, for example, response selection, motor preparation, and finally motor execution (see Thorpe, Fize, & Marlot, 1996), the fact that people were able to refresh at an SOA of 100 ms does not allow an inference of a very precise minimal refreshing duration. However, other research suggests that refreshing a single WM representation requires approximately 50 ms (Camos & Barrouillet, 2014; Vergauwe et al., 2014; Vergauwe & Cowan, 2014; see also Lemaire, Pageot, Plancher, & Portrat, 2018). In a way, then, it is surprising that recent studies on refreshing provide unnecessarily long refreshing times of, for example, 500 ms (Souza et al., 2015) or 1000 ms (Vergauwe & Langerock, 2017). It should be acknowledged, however, that refreshing time is probably dependent on contextual factors. For example, it is reasonable that increasing the number of memoranda results in longer refreshing times. On a speculative nature, people might have to search the WM space in order to arrive at the to-be-refreshed item, analogous to the process of working through the objects in a visual search array (e.g., Luck & Hillyard, 1994; Treisman, 1982; Wolfe, Cave, & Franzel, 1989; Wolfe, 1994; Wolfe, 1998). As a consequence, the more items are in WM (equivalent to set size in visual search), the longer it might take to find the to-be-refreshed representation. Thus, it might be possible that the amount of time for refreshing was higher in the study by Vergauwe and Langerock (2017), where four instead of two memoranda were used. Moreover, refreshing times might be different for various categories and features of memoranda (e.g., alphanumeric characters, pictures, unfamiliar symbols, colors, orientation, valence) and even sensory modalities (e.g., M. K. Johnson et al., 2005; Nees, Corrini, Leong, & Harris, 2017).

In the second experiment, the main objective was to test whether WM consolidation and refreshing rely upon the same resource or separate resources. To be able to investigate this, we established a new paradigm (see Figure 4) that was a mixture of the skeletal RSVP (AB) paradigm and our modified instructed-refreshing paradigm (cf. Vergauwe & Langerock, 2017), with the feasibility of the latter being demonstrated in the first experiment. Participants first saw and had to memorize two letters (memoranda) and subsequently also a picture, which served as T1 in the sense of the classical AB paradigm. Crucially, the refreshing cue was then either presented during the AB (at a PCOA of 300 ms) or afterwards (at a PCOA of 1200 ms). We inferred the magnitude and temporal dynamics of the AB from a separate block of trials in which the AB, elicited by the T1 pictures, was explicitly tested (see Figure 3). Lastly, a probe letter was shown and participants had to decide as quickly as possible whether the probe matched one of the previously presented memoranda or not. Our reasoning was that if certain attentional resources are used for the consolidation of a stimulus, the ability to refresh items in WM at the same time should be reduced if this process of refreshing requires the same attentional resources as the consolidation process. If, however, refreshing relies on different kinds of resources, the ability to refresh information in WM should be independent of and, thus, unaffected by concurrent consolidation processes. In either case, refreshing should be possible outside of the AB window (at a PCOA of 1200 ms). Unfortunately, we did not observe a refresh benefit after the AB, which left us unable to properly test the two competing hypotheses (see Figure 5).

To probe the causes of this surprising finding of a lacking refresh benefit at a PCOA of 1200, we conducted several exploratory analyses. In a first step, we tested whether our results differed according to whether we probed the left or right memorandum. We observed a clear difference in overall RTs between the two locations, with shorter RTs for the left compared to the right probed letter. However, the location of the probed letter did not interact with any main or interaction effect, suggesting that the refresh benefits at the two PCOAs were not much different between the two locations. Hence, the differentiation between the left and right probed letters did not provide any hints about the cause of the lacking refresh benefit at 1200 ms.

This leftward bias is actually quite common and is observed in many different disciplines and paradigms. For example, when viewing scenes, people tend to look more at the left side and the initial saccade is usually directed to the left (Dickinson & Intraub, 2009; Foulsham et al., 2013). Moreover, in a line bisection task, healthy participants tend to bisect the line slightly to the left from the central point, which is called pseudoneglect due to its resemblance to the neurological condition of hemineglect (Bowers & Heilman, 1980). Also, in visual search paradigms people usually start searching for the target on the left side of the display (Williams & Reingold, 2001; Zelinsky, 1996). It has been argued that the leftward bias in all these phenomena might be due to an attentional bias resulting from a right hemisphere dominance (Corbetta & Shulman, 2002; Mesulam, 1999). Furthermore, since we used letters as memoranda, it is reasonable to assume that the well-learned skill of reading had an impact on the leftward bias. That is, for most written languages reading starts at the left side of a document and progresses to the right for each line of text. Something similar might have happened in our design. Obviously, the location and the nature of the memoranda are important factors in refreshing experiments. The implication for future experiments is that these factors need to be taken into account when designing the study; otherwise, it is possible that power is undermined. For example, in the future, memoranda could be presented at the top and at the bottom of the display. Alternatively, the use of less discrete memoranda that are not strictly bound to certain locations might turn out to be advantageous. A good example of that suggestion would be the study by Souza, Rerko, and Oberauer (2015) in which a color wheel was used. Moreover, since it is likely that the highly automatized skill of reading from the left to the right exerts an influence if alphanumeric characters are used, it might be worthwhile to get rid of alphanumeric characters overall and use different stimulus categories instead (e.g., pictures or unfamiliar symbols).

In a second step, we only considered participants that were in the highest tertile in terms of the magnitude of the refresh benefit at 1200 ms. These participants had a large refresh benefit at 1200 ms, allowing us to probe our competing hypotheses. However, although the evidence marginally tended to favor the hypothesis that consolidation and refreshing do not rely upon the same resource, extreme caution should be exercised. For one, the evidence was very weak, almost to the level of not favoring any hypothesis. Additionally, we should remember that we only considered a well-chosen subsample that only consisted of ten participants, thus, undermining the generalizability of any findings.

Finally, in the last step of data exploration we tested whether the magnitude of the refresh benefit at 1200 ms correlated with WM capacity. Admittedly, the task in the second experiment was quite complex. Hence, we reasoned that some people might have handled the task more easily, maybe due to a high WM capacity. To this end, we calculated a tentative index of WM capacity based on the performance measures in the two blocks. However, the evidence indicated that our index of WM was probably not related to the refresh benefit at

1200 ms. Of course, we do not know how a proper index of WM would have correlated with the magnitude of the refresh benefit. It might be worthwhile for future studies to obtain a reliable and valid measure of WM capacity for each participant. It would certainly be interesting to investigate whether and to what extent WM capacity predicts the magnitude of the refresh benefit on an individual basis.

Still, we do not know why, on average, participants did not exhibit a refresh benefit after consolidation was completed (i.e., at a PCOA of 1200 ms). It has been argued that items in WM can be in two different states, namely an active and an accessory or passive state (e.g., Oberauer, 2002; Olivers et al., 2011). Active items are relevant for the current task and are therefore in the focus of central executive processes. Accessory items, in contrast, are merely stored for later use. Importantly, while active items are readily accessible, it takes time and effort to bring accessory items into the focus of attention. It is possible that this two-state account of WM partly explains the lack of a refresh benefit in the combination block of the second experiment. Specifically, when memoranda are presented, they are processed and consolidated, probably receiving an active status in WM due to their significance for the task at hand. However, this privileged active status of the WM representations of the memoranda might be overwritten once the T1 picture is presented, putting the letters in an accessory state. At this point, observers experience a switch in tasks and a rather heavy change in stimulus category (i.e., from simple letters to fine details of a picture). The most severe switch, however, would be to direct attention to the external world when viewing the T1 picture and then quickly focus attention on internal WM representations when the refreshing cue appears. It is known that these switches are associated with time costs (e.g., Kiesel et al., 2010; Monsell, 2003; Oberauer, 2002; Rogers & Monsell, 1995; Rubinstein, Meyer, & Evans, 2001; Visser et al., 1999). Thus, when the refreshing cue arrives, another switch back to the memoranda is required, which might not be possible that easily and quickly. Hence, this could explain the almost absence of a refreshing benefit.

An interesting way to inquire into this possibility is to replace the T1 picture with another set of letters. Participants would have the task to remember both the initial memoranda and the T1 letter set, in the fashion of Oberauer's (2002) study in which some items had to be maintained for further processing and others only more passively for a memory test. In that way, there would not be categorically distinct stimuli. Additionally, it would be much easier to comprehend the task requirements. The refreshing cue that follows the T1 letter at different SOAs would only refer to the first set of letters. If the costs of task switching and differential stimulus categories contributed to the lack of a refresh benefit in our study, we would expect to observe a refresh benefit in the just suggested paradigm. One complication of using letters as T1 might be that the resulting AB would probably be short-lived. This was one of the reasons why we used pictures instead of alphanumeric characters as T1 in the first place.

It is also possible that the nature of the T1 task prevented refreshing. In particular, participants were confronted with the task to memorize the T1 pictures in as much detail as possible. This is quite challenging in light of the fact that the pictures were only shown for 50 ms. Conceivably, participants might have prioritized this task over refreshing, which might have had the consequence that the pictures remained in the focus of attention for a prolonged time in order to extract as many details as possible, even during the presentation of the refreshing cues. Indeed, the overall accuracies for the T1 picture recognition were rather high (see Figure 7). As a result, the limited availability of attention at the probe recognition test might have left the old/new judgement be based on the automatic process of familiarity rather than intentional and active recollection (e.g., Jacoby, 1991; Mandler, 1980; Yonelinas, 2002).

Conclusions

The starting point of the current study was the identification of refreshing as a mechanism to maintain information in WM (e.g., Barrouillet et al., 2004; Barrouillet et al., 2007; Camos & Barrouillet, 2014; Vergauwe et al., 2014; Vergauwe & Langerock, 2017). Presumably, refreshing works by bringing WM representations into the focus of attention, thus, increasing their activation (Barrouillet et al., 2007; M. K. Johnson et al., 2002). It was shown that the local effects of refreshing are increased accessibility, which can be observed in shorter RTs (Vergauwe & Langerock, 2017). However, little is known about how refreshing relates to and interacts with other WM processes (e.g., consolidation).

Using a simplified version of the instructed-refreshing paradigm by Vergauwe and Langerock (2017) in our first experiment, we replicated their finding of a benefit for just refreshed WM representations, in terms of a shorter RT compared to not just refreshed items. Thus, we demonstrated the feasibility of this simpler paradigm, containing only two memoranda presented simultaneously and only one refreshing cue.

In the second experiment, we used a mixture of the skeletal AB paradigm (e.g., Nieuwenstein et al., 2009; Ward et al., 1997) and our simplified instructed-refreshing paradigm (cf. Vergauwe & Langerock, 2017). With that, we could combine the WM processes of consolidation and refreshing within a single dual-task paradigm. Critically, we manipulated whether participants had to refresh letters during or after the consolidation of a picture (i.e., during the AB elicited by the picture). We reasoned that the ability to refresh should be reduced during consolidation, compared to after consolidation, if consolidation and refreshing utilize the same resource. If, however, consolidation and refreshing use separate resources, we would expect to see no reduction in the refresh benefit. Regrettably, we were not able to probe these competing hypotheses because participants did not even display a refresh benefit after consolidation.

We speculate that the lacking refresh benefit after consolidation might be due to several reasons. To point out a few suggestions, it is conceivable that the presentation and consolidation of the picture might have pushed the memoranda into a passive state in WM (Oberauer, 2002; Olivers et al., 2011). Bringing memoranda back into an active state through instructed refreshing might have proven rather difficult, given the strength of the picture. Alternatively, the task for T1 itself might have largely prevented refreshing. Specifically, the task of memorizing the T1 pictures in detail might have been very challenging, given the short presentation duration of only 50 ms. Conceivably, the pictures might have remained in the focus of attention, having the consequence that the probe recognition task might have been based on familiarity rather than recollection (e.g., Jacoby, 1991; Mandler, 1980; Yonelinas, 2002).

In summary, we lay the ground for research on the interdependence of WM-related processes by devising a new dual-task paradigm based on the skeletal AB paradigm (e.g., Nieuwenstein et al., 2009; Ward et al., 1997) and our modified version of the instructed-refreshing paradigm by Vergauwe and Langerock (2017). This new paradigm, however, needs to be developed further. Future research could incorporate some of our suggestions for further improvements.

References

Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Clarendon Press.

Baddeley, A. D. (2003a). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*(10), 829-839.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (pp. 47-89). New York: NY: Academic Press.

Baddeley, A. D. (2003b). Working memory and language: An overview. *Journal of Communication Disorders, 36*(3), 189-208.

Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*(1), 158-173.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods, 37*(3), 379-384.

Barrouillet, P. (1996). Transitive inferences from set-inclusion relations and working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(6), 1408-1422.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General, 133*(1), 83-100.

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(3), 570-585.

Barrouillet, P., & Camos, V. (2015). *Working memory: Loss and reconstruction*. Hove, UK: Psychology Press.

Barrouillet, P., & Camos, V. (2012). As time goes by: Temporal constraints in working memory. *Curr Dir Psychol Sci, 21*(6), 413-419.

Barrouillet, P., Uittenhove, K., Lucidi, A., & Langerock, N. (2017). On the sources of forgetting in working memory: The test of competing hypotheses. *Quarterly Journal of Experimental Psychology, 00*(0), 1-20.

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General, 132*(1), 71-92.

Bowers, D., & Heilman, K. M. (1980). Pseudoneglect: Effects of hemispace on a tactile line bisection task. *Neuropsychologia, 18*(4-5), 491-498.

Bowman, H., & Wyble, B. (2007). The simultaneous type, serial token model of temporal attention and working memory. *Psychological Review, 114*(1), 38-70.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences of the United States of America, 105*(38), 14325-14329.

Breitmeyer, B. G., & Öğmen, H. (2006). *Visual masking: Time slices through conscious and unconscious vision* (2nd ed.). Oxford, UK: Oxford University Press.

Broadbent, D. E., & Broadbent, M. H. P. (1987). From detection to identification: Response to multiple targets in rapid serial visual presentation. *Perception & Psychophysics, 42*(2), 105-113.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *The Quarterly Journal of Experimental Psychology, 10*, 12-21.

Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015). Intrinsic and extrinsic effects on image memorability. *Vision Research, 116*, 165-178.

Camos, V., Lagner, P., & Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language, 61*(3), 457-469.

Camos, V., & Barrouillet, P. (2014). Attentional and non-attentional systems in the maintenance of verbal information in working memory: The executive and phonological loops. *Frontiers in Human Neuroscience, 8*, 1-11.

Chun, M. M. (1997a). Temporal binding errors are redistributed by the attentional blink. *Perception & Psychophysics, 59*(8), 1191-1199.

Chun, M. M. (1997b). Types and tokens in visual processing: A double dissociation between the attentional blink and repetition blindness. *Journal of Experimental Psychology: Human Perception and Performance, 23*(3), 738-755.

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance, 21*(1), 109-127.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55*(4), 429-432.

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*(2), 163-183.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience, 3*(3), 215-229.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87-185.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science, 19*(1), 51-57.

Cowan, N., Wood, N. L., Wood, P. K., Keller, T. A., Nugent, L. D., & Keller, C. V. (1998). Two separate verbal processing rates contributing to short-term memory span. *Journal of Experimental Psychology: General, 127*(2), 141-160.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450-466.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*(4), 422-433.

DeStefano, D., & LeFevre, J. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology, 16*(3), 353-386.

Di Lollo, V., Kawahara, J., Ghorashi, S. M. S., & Enns, J. T. (2005). The attentional blink: Resource depletion or temporary loss of control? *Psychological Research, 69*(3), 191-200.

Dickinson, C. A., & Intraub, H. (2009). Spatial asymmetries in viewing and remembering scenes: Consequences of an attentional bias? *Attention, Perception, & Psychophysics, 71*(6), 1251-1262.

Duncan, J., Ward, R., & Shapiro, K. L. (1994). Direct measurement of attentional dwell time in human vision. *Nature, 369*(6478), 313-315.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1*(2), 170-177.

Dux, P. E., & Marois, R. (2009). The attentional blink: A review of data and theory. *Attention, Perception, & Psychophysics, 71*(8), 1683-1700.

Ecker, U. K. H., Lewandowsky, S., & Oberauer, K. (2014). Removal of information from working memory: A specific updating process. *Journal of Memory and Language, 74*, 77-90.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake, & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102-134). Cambridge, England: Cambridge University Press.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(5), 972-992.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*(3), 309-331.

Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2007). Masking disrupts reentrant processing in human visual cortex. *Journal of Cognitive Neuroscience, 19*(9), 1488-1497.

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191.

Foulsham, T., Gray, A., Nasiopoulos, E., & Kingstone, A. (2013). Leftward biases in picture scanning and line bisection: A gaze-contingent window study. *Vision Research, 78*, 14-25.

Garavan, H. (1998). Serial attention within working memory. *Memory & Cognition, 26*(2), 263-276.

Giesbrecht, B., & Di Lollo, V. (1998). Beyond the attentional blink: Visual masking by object substitution. *Journal of Experimental Psychology: Human Perception and Performance, 24*(5), 1454-1466.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research, 42*(3), 237-288.

Goldman-Rakic, P. S. (1992). Working memory and the mind. *Scientific American, 267*(3), 110-117.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*, 95-112.

Haberlandt, K. (1997). *Cognitive psychology*. Boston, MA: Allyn and Bacon.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences*, *21*(6), 803-864.

Harrison, T. L., Shipstead, Z., & Engle, R. W. (2015). Why is working memory capacity related to matrix reasoning tasks? *Memory & Cognition, 43*(3), 389-396.

Heathcote, A., Popiel, S. J., & Mewhort, D. J. (1991). Analysis of response time distributions: An example using the stroop task. *Psychological Bulletin, 109*(2), 340-347.

Henson, R. N. A. (1999). Positional information in short-term memory: Relative or absolute? *Memory & Cognition, 27*(5), 915-927.

Hommel, B., & Akyürek, E. (2005). Lag-1 sparing in the attentional blink: Benefits and costs of integrating two events into a single episode. *Quarterly Journal of Experimental Psychology: Section A, 58*(8), 1415-1433.

Hudjetz, A., & Oberauer, K. (2007). The effects of processing time and processing rate on forgetting in working memory: Testing four models of the complex span paradigm. *Memory & Cognition, 35*(7), 1675-1684.

Huynh, H., & Feldt, L. S. (1976). Estimation of the box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics, 1*(1), 69-82.

Isola, P., Parikh, D., Torralba, A., & Oliva, A. (2011). (2011). Understanding the intrinsic memorability of images. Paper presented at the *Advances in Neural Information Processing Systems,* 2429-2437.

Isola, P., Xiao, J., Parikh, D., Torralba, A., & Oliva, A. (2014). What makes a photograph memorable? *IEEE Transactions on Pattern Analysis and Machine Intelligence, 36*(7), 1469-1482.

Isola, P., Xiao, J., Torralba, A., & Oliva, A. (2011). (2011). What makes an image memorable? Paper presented at the *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference On,* Providence, RI. 145-152.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513-541.

JASP Team. (2018). *JASP (Version 0.9)[computer software]*

Johnson, M. K., Reeder, J. A., Raye, C. L., & Mitchell, K. J. (2002). Second thoughts versus second looks: An age-related deficit in reflectively refreshing just-activated information. *Psychological Science, 13*(1), 64-67.

Johnson, M. K., Raye, C. L., Mitchell, K. J., Greene, E. J., Cunningham, W. A., & Sanislow, C. A. (2005). Using fMRI to investigate a component process of reflection: Prefrontal correlates of refreshing a just-activated representation. *Cognitive, Affective & Behavioral Neuroscience, 5*(3), 339-361.

Johnson, M. R., Higgins, J. A., Norman, K. A., Sederberg, P. B., Smith, T. A., & Johnson, M. K. (2013). Foraging for thought: An inhibition-of-return-like effect resulting from directing attention within working memory. *Psychological Science, 24*(7), 1104-1112.

Jolicœur, P., & Dell'Acqua, R. (1998). The demonstration of short-term consolidation. *Cognitive Psychology, 36*(2), 138-202.

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

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*(2), 189-217.

Kanwisher, N. G. (1987). Repetition blindness: Type recognition without token individuation. *Cognition, 27*(2), 117-143.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773-795.

Kawahara, J., Enns, J. T., & Di Lollo, V. (2006). The attentional blink is not a unitary phenomenon. *Psychological Research, 70*(6), 405-413.

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin, 136*(5), 849-874.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14*(4), 389-433.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4,* 863.

Lemaire, B., Pageot, A., Plancher, G., & Portrat, S. (2018). What is the time course of working memory attentional refreshing? *Psychonomic Bulletin & Review, 25*(1), 370-385.

Levi, D. M., McGraw, P. V., & Klein, S. A. (2000). Vernier and contrast discrimination in central and peripheral vision. *Vision Research, 40*(8), 973-988.

Lewandowsky, S., Duncan, M., & Brown, G. D. A. (2004). Time does not cause forgetting in short-term serial recall. *Psychonomic Bulletin & Review, 11*(5), 771-790.

Lewandowsky, S., Oberauer, K., & Brown, G. D. A. (2009). No temporal decay in verbal short-term memory. *Trends in Cognitive Sciences, 13*(3), 120-126.

Linde, M., Nieuwenstein, M. R., & van Rijn, H. (2016). *Dwelling on memorability: Effects of long-term picture memorability on the attentional blink.* Unpublished manuscript.

Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology, 31*(3), 291-308.

Lunney, G. H. (1970). Using analysis of variance with a dichotomous dependent variable: An empirical study. *Journal of Educational Measurement, 7*(4), 263-269.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87*(3), 252-271.

Mathôt, S. (2017, 15 May). Bayes like a baws: Interpreting bayesian repeated measures in JASP. Retrieved from https://www.cogsci.nl/blog/interpreting-bayesian-repeated-measures-in-jasp

Mesulam, M. M. (1999). Spatial attention and neglect: Parietal, frontal and cingulate contributions to the mental representation and attentional targeting of salient extrapersonal events. *Philosophical Transactions: Biological Sciences, 354*(1387), 1325-1346.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7*(3), 134-140.

Morey, R. D., & Rouder, J. N. (2015). *BayesFactor: Computation of bayes factors for common designs* (R package version 0.9.12-2 ed.)

Most, S. B., Chun, M. M., Widders, D. M., & Zald, D. H. (2005). Attentional rubbernecking: Cognitive control and personality in emotion-induced blindness. *Psychonomic Bulletin & Review, 12*(4), 654-661.

Most, S. B., Smith, S. D., Cooter, A. B., Levy, B. N., & Zald, D. H. (2007). The naked truth: Positive, arousing distractors impair rapid target perception. *Cognition and Emotion, 21*(5), 964-981.

Naveh-Benjamin, M., & Jonides, J. (1984). Maintenance rehearsal: A two-component analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(3), 369-385.

Nee, D. E., & Jonides, J. (2008). Neural correlates of access to short-term memory. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 105*(37), 14228-14233.

Nees, M. A., Corrini, E., Leong, P., & Harris, J. (2017). Maintenance of memory for melodies: Articulation or attentional refreshing? *Psychonomic Bulletin & Review, 24*(6), 1964-1970.

Nieuwenstein, M. R. (2006). Top-down controlled, delayed selection in the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance, 32*(4), 973-985.

Nieuwenstein, M. R., Chun, M. M., van der Lubbe, R. H. J., & Hooge, I. T. C. (2005). Delayed attentional engagement in the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance, 31*(6), 1463-1475.

Nieuwenstein, M. R., & Kromm, M. (2017). Visual memory: The price of encoding details. *Paper Presented at the 58th Annual Meeting of the Psychonomic Society, Vancouver, Canada,* Nieuwenstein, M. R., & Potter, M. C. (2006). Temporal limits of selection and memory encoding: A comparison of whole versus partial report in rapid serial visual presentation. *Psychological Science, 17*(6), 471-475.

Nieuwenstein, M. R., Potter, M. C., & Theeuwes, J. (2009). Unmasking the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance, 35*(1), 159-169.

Nieuwenstein, M. R., Van der Burg, E., Theeuwes, J., Wyble, B., & Potter, M. (2009). Temporal constraints on conscious vision: On the ubiquitous nature of the attentional blink. *Journal of Vision, 9*(9), 1-14.

Nieuwenstein, M. R., & Wyble, B. (2014). Beyond a mask and against the bottleneck: Retroactive dual-task interference during working memory consolidation of a masked visual target. *Journal of Experimental Psychology: General, 143*(3), 1409-1427.

Oberauer, K. (2001). Removing irrelevant information from working memory: A cognitive aging study with the modified sternberg task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(4), 948-957.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 411-421.

Oberauer, K., Farrell, S., Jarrold, C., Pasiecznik, K., & Greaves, M. (2012). Interference between maintenance and processing in working memory: The effect of item–distractor similarity in complex span. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(3), 665-685.

Oberauer, K., & Hein, L. (2012). Attention to information in working memory. *Curr Dir Psychol Sci, 21*(3), 164-169.

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language, 55*(4), 601-626.

Oberauer, K., & Lewandowsky, S. (2008). Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference? *Psychological Review, 115*(3), 544-576.

Oberauer, K., & Lewandowsky, S. (2014). Further evidence against decay in working memory. *Journal of Memory and Language, 73*, 15-30.

Oberauer, K., & Lewandowsky, S. (2016). Control of information in working memory: Encoding and removal of distractors in the complex-span paradigm. *Cognition, 156*, 106-128.

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review, 19*(5), 779-819.

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods, 8*(4), 434-447.

Olivers, C. N. L., & Nieuwenhuis, S. (2006). The beneficial effects of additional task load, positive affect, and instruction on the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance, 32*(2), 364-379.

Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences, 15*(7), 327-334.

Olivers, C. N. L., Van der Stigchel, S., & Hulleman, J. (2007). Spreading the sparing: Against a limited-capacity account of the attentional blink. *Psychological Research, 71*(2), 126-139.

Öztekin, I., Davachi, L., & McElree, B. (2010). Are representations in working memory distinct from representations in long-term memory? neural evidence in support of a single store. *Psychological Science*, *21*(8), 1123-1133.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116*(2), 220-244.

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*(3), 193-198.

Potter, M. C., Chun, M. M., Banks, B. S., & Muckenhoupt, M. (1998). Two attentional deficits in serial target search: The visual attentional blink and an amodal task-switch deficit. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(4), 979-992.

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*(3), 510-532.

Raye, C. L., Johnson, M. K., Mitchell, K. J., Greene, E. J., & Johnson, M. R. (2007). Refreshing: A minimal executive function. *Cortex, 43*(1), 135-145.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance, 18*(3), 849-860.

Rerko, L., Oberauer, K., & Lin, H. (2014). Spatial transposition gradients in visual working memory. *The Quarterly Journal of Experimental Psychology, 67*(1), 3-15.

Ricker, T. J., Nieuwenstein, M. R., Bayliss, D. M., & Barrouillet, P. (2018). Working memory consolidation: Insights from studies on attention and working memory. *Annals of the New York Academy of Sciences,* Ricker, T. J., Vergauwe, E., & Cowan, N. (2016). Decay theory of immediate memory: From brown (1958) to today (2014). *The Quarterly Journal of Experimental Psychology, 69*(10), 1969-1995.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictible switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124*(2), 207-231.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default bayes factors for ANOVA designs. *British Journal of Mathematical and Statistical Psychology, 56*(5), 356-374.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225-237.

Rubinstein, J. S., Meyer, D. E., & Evans, J. E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance, 27*(4), 763-797.

Schönbrodt, F. D., Wagenmakers, E., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with bayes factors: Efficiently testing mean differences. *Psychological Methods, 22*(2), 322-339.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*(3/4), 591-611.

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2018). *Afex: Analysis of factorial experiments* (R package version 0.20-2 ed.)

Souza, A. S., Rerko, L., & Oberauer, K. (2014). Unloading and reloading working memory: Attending to one item frees capacity. *Journal of Experimental Psychology: Human Perception and Performance, 40*(3), 1237-1256.

Souza, A. S., Rerko, L., & Oberauer, K. (2015). Refreshing memory traces: Thinking of an item improves retrieval from visual working memory. *Annals of the New York Academy of Sciences, 1339,* 20-31.

Standing, L. (1973). Learning 10000 pictures. *The Quarterly Journal of Experimental Psychology, 25*(2), 207-222.

Süß, H., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability--and a little bit more. *Intelligence, 30*(3), 261-288.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520-522.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance, 8*(2), 194-214.

Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General, 123*(1), 34-80.

Van Selst, M., & Jolicœur, P. (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 47A*(3), 631-650.

VanRullen, R., & Koch, C. (2003). Is perception discrete or continuous? *Trends in Cognitive Sciences, 7*(5), 207-213.

Vergauwe, E., Camos, V., & Barrouillet, P. (2014). The impact of storage on processing: How is information maintained in working memory? *Journal of Experimental Psychology.Learning, Memory, and Cognition, 40*(4), 1072-1095.

Vergauwe, E., Barrouillet, P., & Camos, V. (2010). Do mental processes share a domain-general resource? *Psychological Science, 21*(3), 384-390.

Vergauwe, E., & Cowan, N. (2014). A common short-term memory retrieval rate may describe many cognitive procedures. *Frontiers in Human Neuroscience, 8,* 126.

Vergauwe, E., & Cowan, N. (2015). Attending to items in working memory: Evidence that refreshing and memory search are closely related. *Psychonomic Bulletin & Review, 22*(4), 1001-1006.

Vergauwe, E., Hardman, K. O., Rouder, J. N., Roemer, E., McAllaster, S., & Cowan, N. (2016). Searching for serial refreshing in working memory: Using response times to track the content of the focus of attention over time. *Psychonomic Bulletin & Review, 23*(6), 1818-1824.

Vergauwe, E., & Langerock, N. (2017). Attentional refreshing of information in working memory: Increased immediate accessibility of just-refreshed representations. *Journal of Memory and Language, 96*, 23-35.

Visser, T. A. W., Bischof, W. F., & Di Lollo, V. (1999). Attentional switching in spatial and nonspatial domains: Evidence from the attentional blink. *Psychological Bulletin*, *125*(4), 458-469.

Vogel, E. K., & Luck, S. J. (2002). Delayed working memory consolidation during the attentional blink. *Psychonomic Bulletin & Review, 9*(4), 739-743.

Vogel, E. K., Luck, S. J., & Shapiro, K. L. (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance, 24*(6), 1656-1674.

Vul, E., Nieuwenstein, M., & Kanwisher, N. (2008). Temporal selection is suppressed, delayed, and diffused during the attentional blink. *Psychological Science, 19*(1), 55-61.

Wagenmakers, E. (2007). A practical solution to the pervasive problems ofp values. *Psychonomic Bulletin & Review, 14*(5), 779-804.

Wagenmakers, E., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., . . . Morey, R. D. (2018). Bayesian inference for psychology. part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review, 25*(1), 35-57.

Ward, R., Duncan, J., & Shapiro, K. (1997). Effects of similarity, difficulty, and nontarget presentation on the time course of visual attention. *Perception & Psychophysics, 59*(4), 593-600.

Weichselgartner, E., & Sperling, G. (1987). Dynamics of automatic and controlled visual attention. *Science, 238*(4828), 778-780.

Williams, D. E., & Reingold, E. M. (2001). Preattentive guidance of eye movements during triple conjunction search tasks: The effects of feature discriminability and saccadic amplitude. *Psychonomic Bulletin & Review, 8*(3), 476-488.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1*(2), 202-238.

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science, 9*(1), 33-39.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 419-433.

Wyble, B., Bowman, H., & Nieuwenstein, M. R. (2009). The attentional blink provides episodic distinctiveness: Sparing at a cost. *Journal of Experimental Psychology: Human Perception and Performance, 35*(3), 787-807.

Wyble, B., Potter, M. C., Bowman, H., & Nieuwenstein, M. (2011). Attentional episodes in visual perception. *Journal of Experimental Psychology: General, 140*(3), 488-505.

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). (2010). Sun database: Large-scale scene recognition from abbey to zoo. Paper presented at the *Computer Vision* *and Pattern Recognition (CVPR), 2010 IEEE Conference On,* San Francisco, CA. 3485-3492.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*(3), 441-517.

Zelinsky, G. J. (1996). Using eye saccades to assess the selectivity of search movements. *Vision Research, 36*(14), 2177-2187.

Zhang, D., Shao, L., Nieuwenstein, M., & Zhou, X. (2008). Top-down control is not lost in the attentional blink: Evidence from intact endogenous cuing. *Experimental Brain Research, 185*(2), 287-295.

Footnotes

[1]The Cohen's *d*, used for the power analysis, is not equivalent to the Cohen's *d* we calculated for our results (Dunlap et al., 1996); moreover, we decided to use the generalized eta squared (η_G^2) instead of the partial eta squared (η_p^2) because it allows better comparisons between uncorrelated and correlated groups designs (Bakeman, 2005).

[2]All analyses were also conducted for data that included this participant. Importantly, statistical conclusions did not change when this participant was included.

[3]The term 'heterogenous' is used to describe the standard errors to point out that they represent the individual unpooled standard error of each mean instead of a homogenous or pooled standard error, which would be equal for each group mean.

[4]All analyses were also conducted for data that included these three participants. Statistical conclusions did not change for digit identification and picture recognition in the AB block and picture recognition in the combination block. Statistical results for probe recognition RTs in the combination block were slightly different. Here, we found moderate evidence against the main effect model of Condition (BF = 0.317), anecdotal evidence against the PCOA main effect model (BF = 0.986), moderate evidence against the combined main effects model of Condition and PCOA (BF = 0.301), and strong evidence against the combined main effects and interaction model (BF = 0.099).

[5]In contrast to our procedure, Mathôt (2017) suggested to divide the posterior probabilities instead of the BFs to calculate a BF_{incl}.

[6]This way of doing the Bayesian dependent-samples one-sided *t*-test coincides with the way this test is performed in JASP (JASP Team, 2018).

[7]We performed a 2 (Block order: AB-combination vs. combination-AB) × 5 (SOA: 300, 700, 1100, 1500, 1900 ms) Bayesian repeated measures ANOVA on digit identification accuracy to determine whether Block order had an effect on accuracy and whether Block

order interacted with SOA. The data were 2.921 times less likely for models that include Block order compared to those that do not (BF_{incl} = 0.342). Furthermore, the data did not favor the inclusion of the interaction between Block order and SOA (BF_{incl} = 0.120). Based on these results, we ignored Block order in further analyses.

[8]We performed a 2 (Block order: AB-combination vs. combination-AB) × 2 (Condition: JR vs. NJR) × 2 (PCOA: 300 vs. 1200 ms) Bayesian repeated measures ANOVA on probe RT to determine whether Block order influenced RT and whether Block order interacted with Condition, PCOA, and the interaction of Condition and PCOA. The data changed the odds against models that include Block order by a factor of 1.711 (BF_{incl} = 0.585). Additionally, the data did not favor the inclusion of the two-way interactions between Block order and Condition (BF_{incl} = 0.310) and Block order and PCOA (BF_{incl} = 0.226) and the three-way interaction of Block order, Condition, and PCOA (BF_{incl} = 0.038). Based on these results, we ignored Block order in further analyses.

Figures

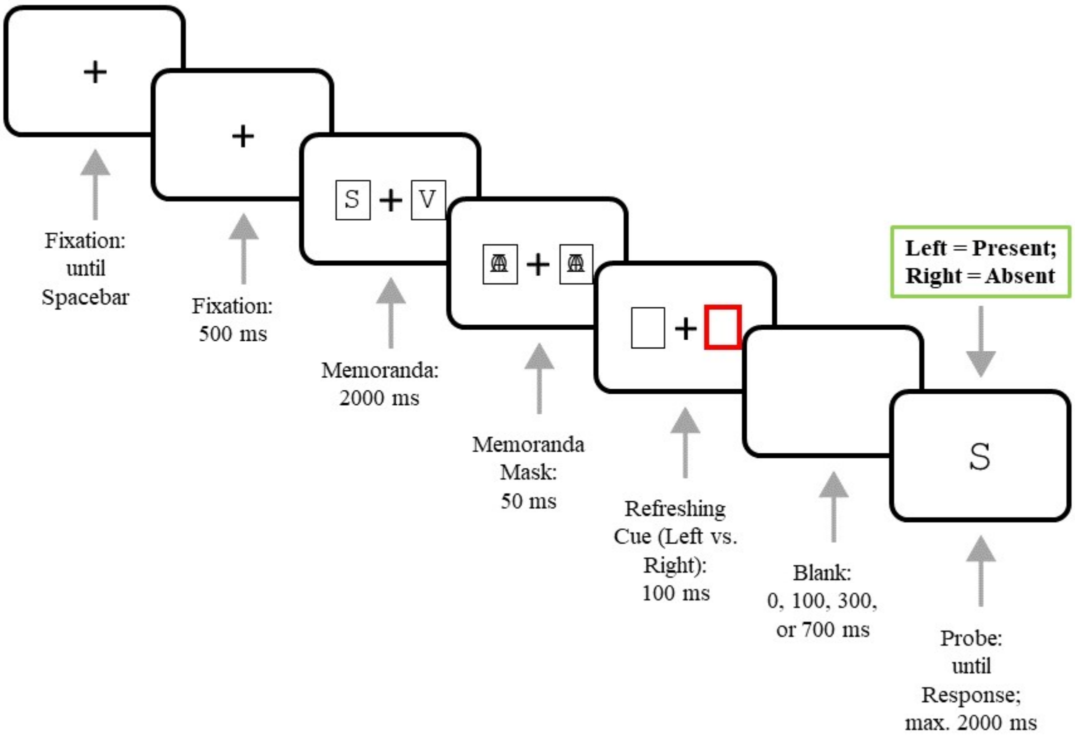

Figure 1. Trial procedure for the first experiment (ms = milliseconds; max. = maximum).

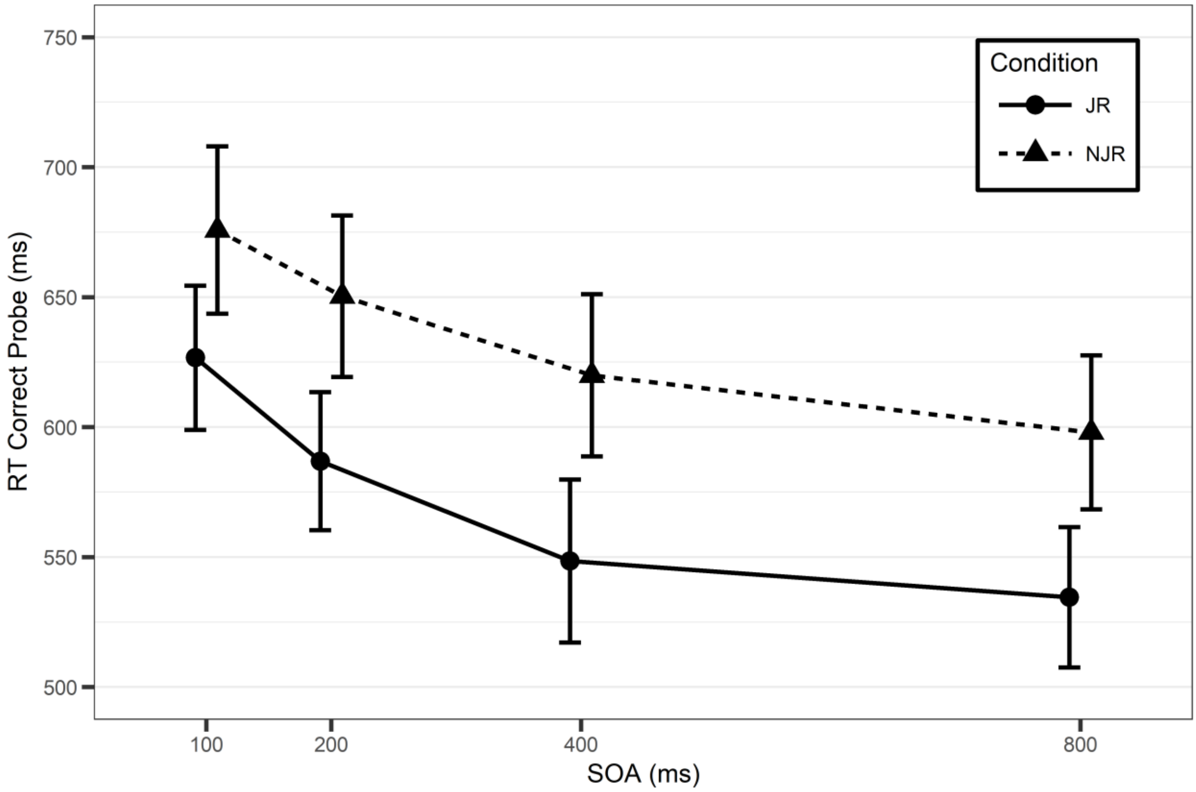

Figure 2. Experiment 1: Mean reaction times for correct responses to the probe at each stimulus onset asynchrony for the two conditions (JR vs. NJR). Error bars represent heterogenous between-subjects standard errors of the mean. Note: ms = milliseconds, SOA = stimulus onset asynchrony between the refreshing cue and the probe, RT = reaction time, JR = just refreshed, NJR = not just refreshed.

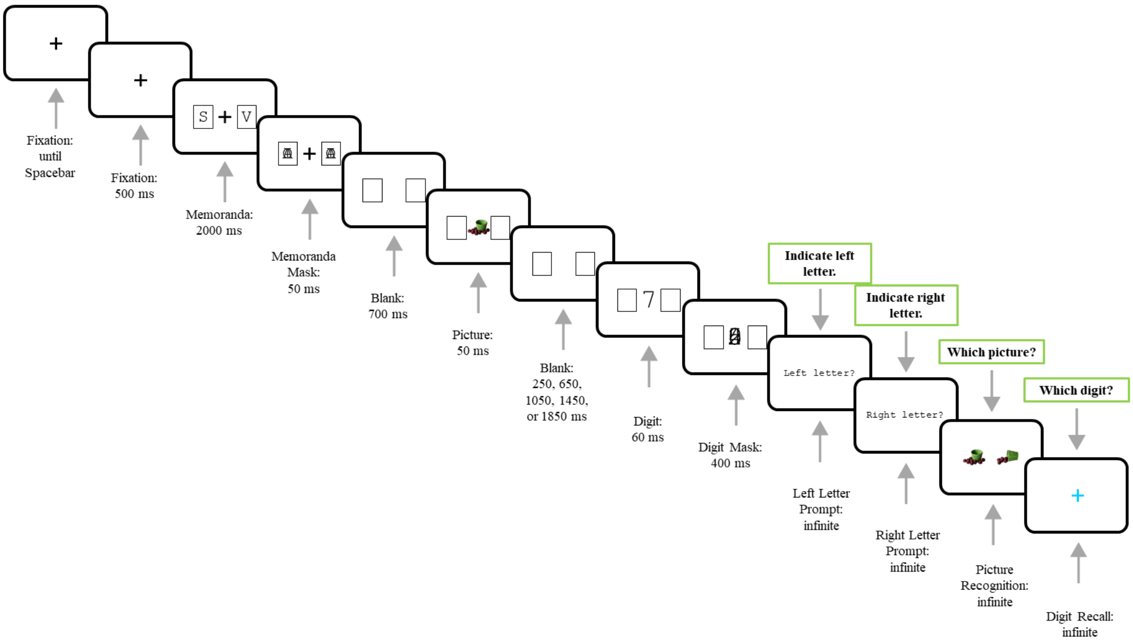

Figure 3. Trial procedure of the attentional blink (AB) block in the second experiment (ms = milliseconds).

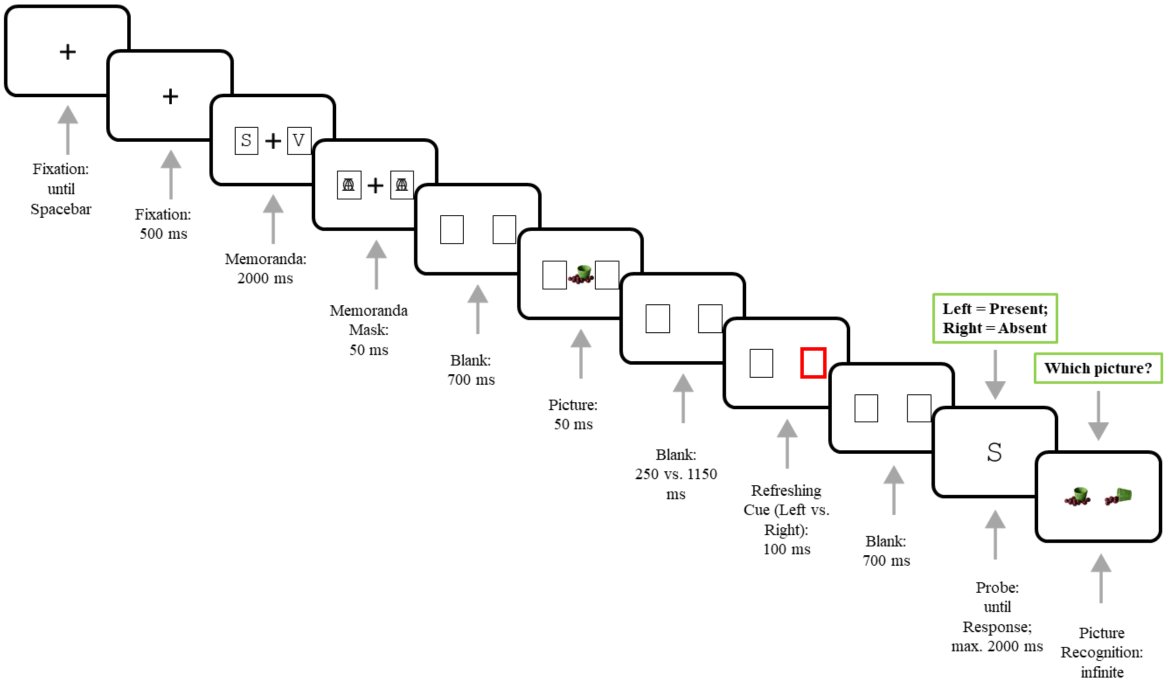

Figure 4. Trial procedure of the combination block in the second experiment (ms = milliseconds; max. = maximum).

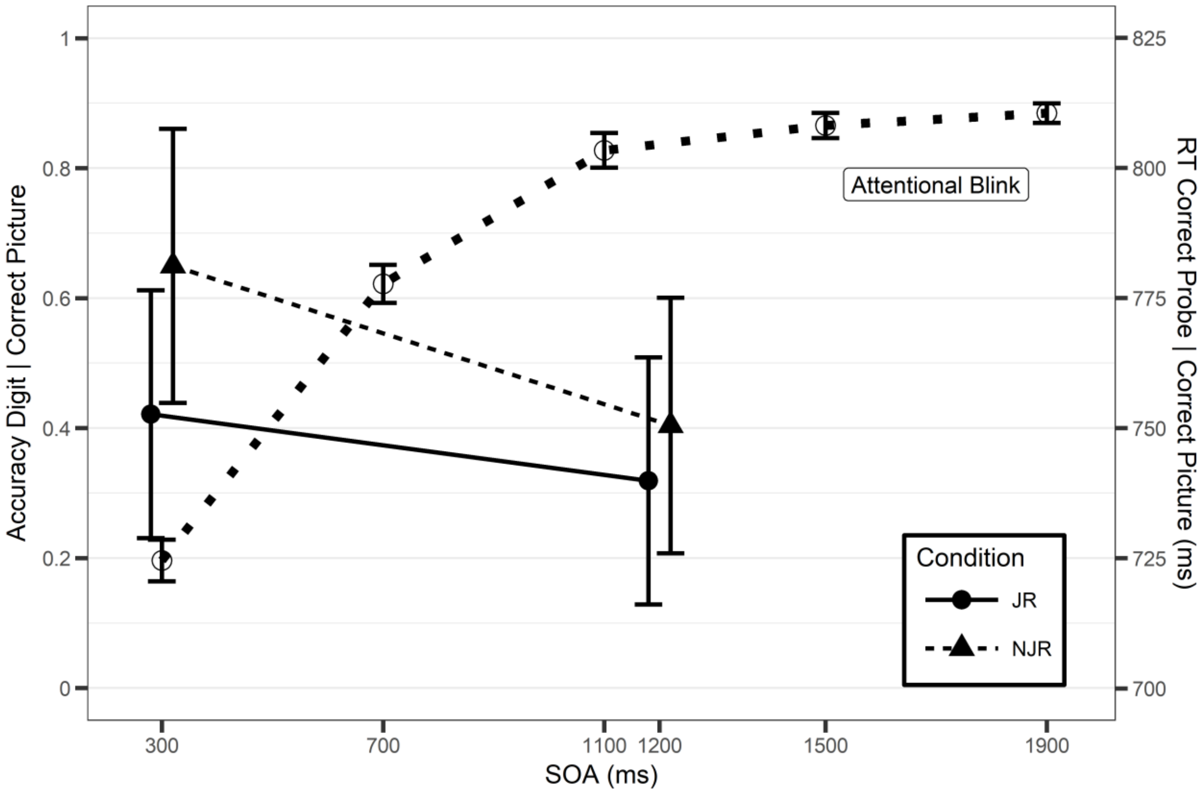

Figure 5. Experiment 2: Mean accuracies for digit identification, for trials in which the picture was recognized correctly, at different stimulus onset asynchronies in the attentional blink (AB) block and mean reaction times for correctly identified probes, for trials in which the picture was recognized correctly, at different picture-cue onset asynchronies (here simply termed stimulus onset asynchrony) in the combination block. The dotted line with empty circles shows the mean accuracies for digit identification in the AB block and refers to the y-axis on the left. The solid and dashed lines (with filled circles and triangles, respectively) show mean reaction times for probe identification in the combination block and refer to the y-axis on the right. Error bars represent heterogenous between-subjects standard errors of the mean. Note: ms = milliseconds, SOA = stimulus onset asynchrony, RT = reaction time, JR = just refreshed, NJR = not just refreshed.

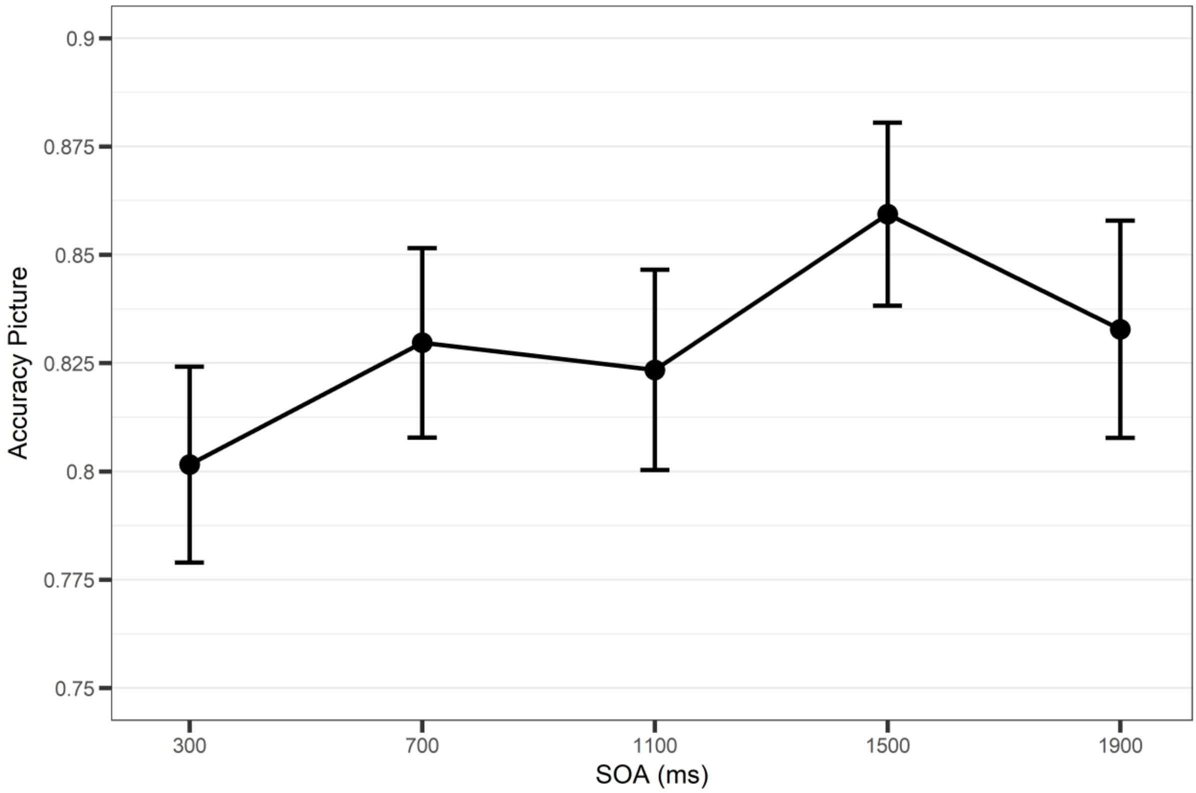

Figure 6. AB block of experiment 2: Mean accuracies for picture recognition at different stimulus onset asynchronies. Error bars represent heterogenous between-subjects standard errors of the mean. Note: ms = milliseconds, SOA = stimulus onset asynchrony.

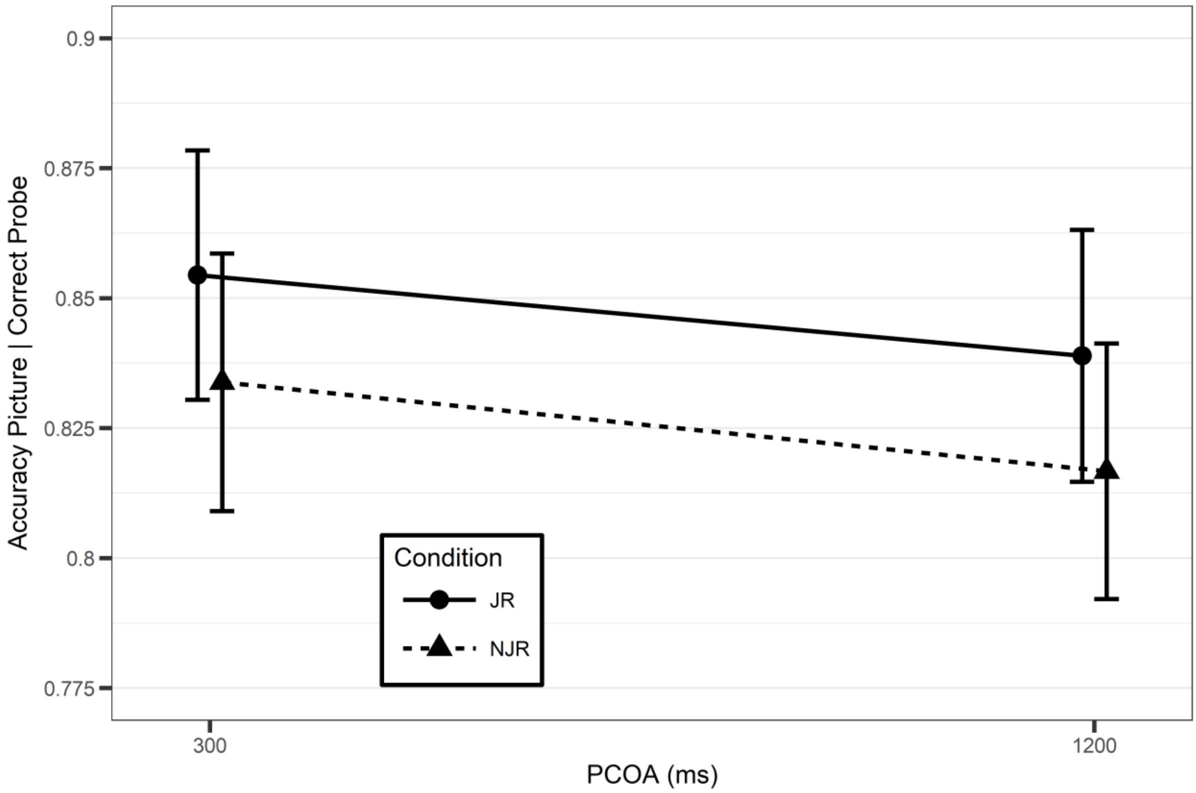

Figure 7. Combination block of experiment 2: Mean accuracies for picture recognition at different picture-cue onset asynchronies for trials in which the probe was correctly recognized. Error bars represent heterogenous between-subjects standard errors of the mean. Note: ms = milliseconds, PCOA = picture-cue onset asynchrony, JR = just refreshed, NJR = not just refreshed.

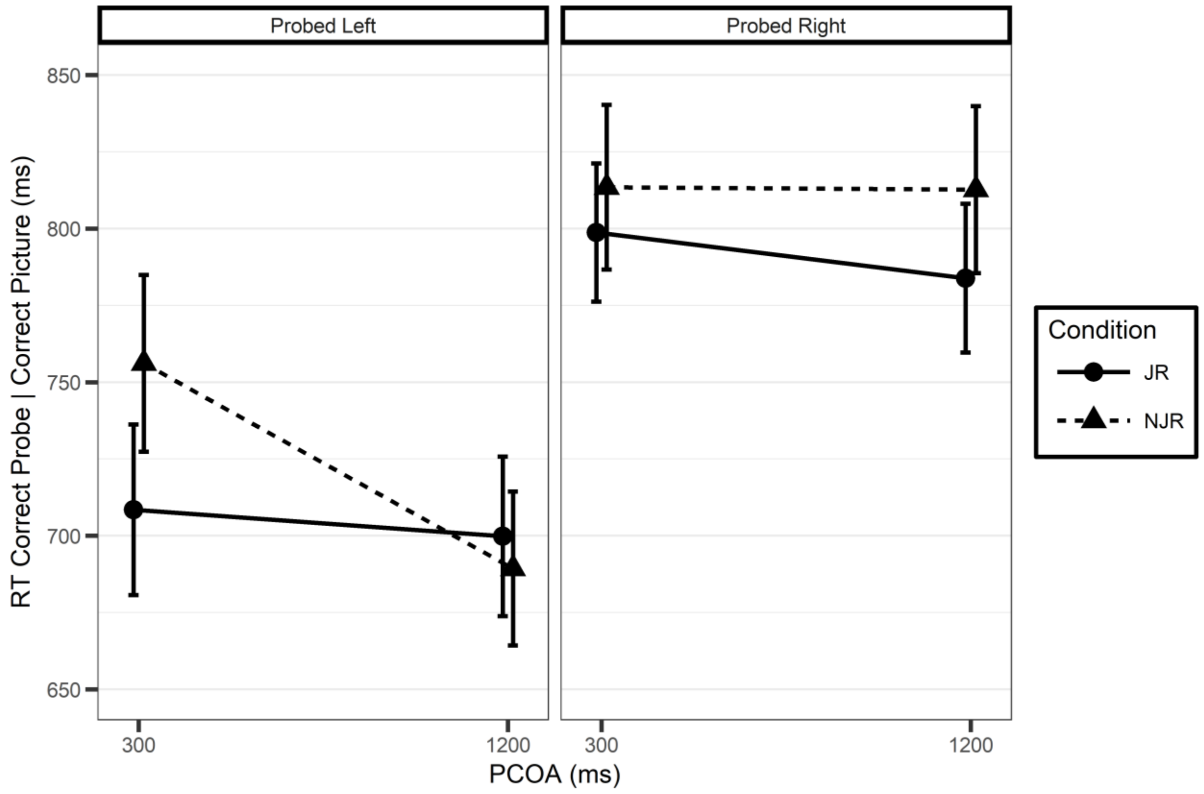

Figure 8. Combination block of experiment 2: Mean reaction times for correct responses to the probe, for trials in which the picture was correctly recognized, at each picture-cue onset asynchrony for the two conditions (JR vs. NJR). The left panel shows mean reaction times for trials in which the left letter was probed, whereas the right panel shows mean reaction times for trials in which the right letter was probed. Error bars represent heterogenous between-subjects standard errors of the mean. Note: ms = milliseconds, PCOA = picture-cue onset asynchrony between the refreshing cue and the probe, RT = reaction time, JR = just refreshed, NJR = not just refreshed.

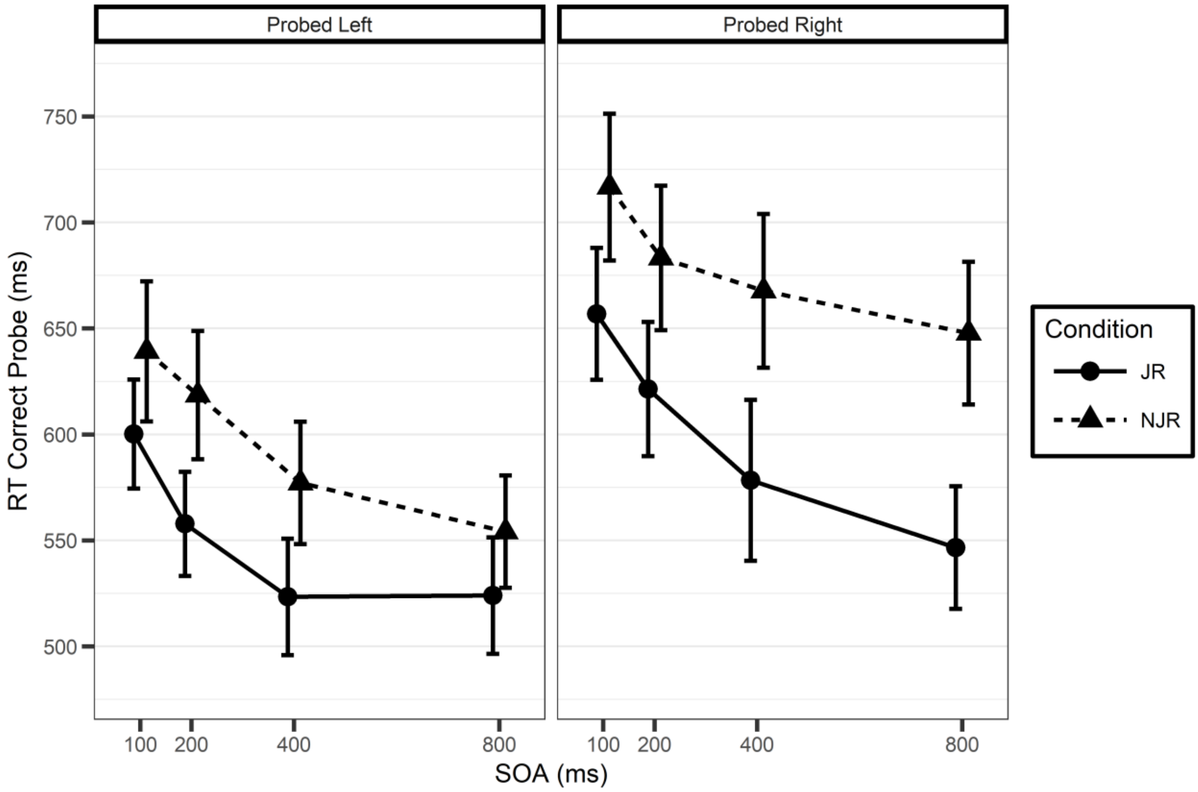

Figure 9. AB block of experiment 1: Mean reaction times for correct responses to the probe at each stimulus onset asynchrony for the two conditions (JR vs. NJR). The left panel shows mean reaction times for trials in which the left letter was probed, whereas the right panel shows mean reaction times for trials in which the right letter was probed. Error bars represent heterogenous between-subjects standard errors of the mean. Note: ms = milliseconds, SOA = stimulus onset asynchrony between the refreshing cue and the probe, RT = reaction time, JR = just refreshed, NJR = not just refreshed.