

Decomposing the n-back task: An individual differences study using the reference-back paradigm

Rachel Rac-Lubashevsky^{a,*}, Yoav Kessler^b

^a Department of Brain and Cognitive Sciences and Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, Israel

^b Department of Psychology and Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, Israel

ARTICLE INFO

Article history:

Received 2 December 2015

Received in revised form

14 June 2016

Accepted 12 July 2016

Available online 15 July 2016

Keywords:

Working memory updating

n-back

Cognitive control

Gating

Reference-back

Individual differences

ABSTRACT

Working memory (WM) has two major functions: *Maintenance*, which is the ability to shield information from being overwritten by irrelevant information, and *updating*, the ability to modify the maintained information when needed. These two conflicting demands are suggested to be controlled by a gating mechanism (for review see O'Reilly (2006)) which enables selective control over updating. Information is robustly maintained in WM when the gate is closed, while opening the gate enables updating. In the present study, we utilized the reference-back paradigm in order to examine their unique contribution to individual differences in n-back, presumably the most widely-used WM updating task. The reference-back is composed of two types of trials: *reference trials* which require both matching (i.e., a same/different judgment) and WM updating, and *comparison trials* which require matching and maintenance. Eighty-eight participants performed the following tasks: 2-choice RT, 1-back, 2-back and the reference-back task. A multiple regression approach was taken in order to explain individual differences in 1-back and 2-back. The reference-back task enabled separating the contribution of the matching decision (difference between mismatch and match), gate-opening (the switch cost in reference trials), gate closing (the switch cost in comparison trials) and WM updating (the difference between reference and comparison trials) to task performance. An intrusion component (WM based proactive interference) was also calculated from 2-back performance. The results indicate that RT in 1-back is mainly predicted by gate opening and by WM updating while 2-back is mainly predicted by gate closing and intrusion. These results confirmed that n-back is not merely an updating task, but also that controlling the contents of WM is the main source of individual differences in the task. The implications for understanding the n-back task and WM updating in general are discussed.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Working memory (WM) is a short-term memory system that supports high-level cognitive processes such as comprehension, reasoning and problem solving (Baddeley, 1992). WM has two general functions: storage (also termed “maintenance”) and processing. The storage function is responsible for keeping relevant information accessible for online processing, such as in complex cognitive tasks. The storage capacity of WM is severely limited, although theorists debate over the exact nature of this limitation: the number of items that could be held simultaneously (e.g., Luck and Vogel, 1997, 2013), or the amount of shared resources that are divided between WM representations (e.g., Bays and Husain, 2008; Ma et al., 2014). Importantly, information held in WM is not only stored passively, but rather can be modified and manipulated.

What makes working memory a *working* system is its processing function, namely the ability to update, manipulate and coordinate information. This ability underlies information processing in complex and novel situations, such as solving a multistage arithmetic problem or planning one's flight schedule for an upcoming conference.

The focus of this study is WM updating, defined as the modification of items held in WM (Kessler and Meiran, 2008; Morris and Jones, 1990; Oberauer, 2001), the addition of new items to WM, and/or the removal of no-longer relevant information from WM (Ecker et al., 2014a, 2014b). While the maintenance state of WM enables it to hold information in a stable manner, impenetrable to irrelevant distraction from the environment, WM also provides the ability to update this information when required. There is an inherent conflict between maintenance and updating as the new information can either be relevant and trigger updating, or irrelevant and hence should be prevented from entering WM. Such conflict requires a control mechanism that regulates the two functions. Computational models of WM (see Chatham and

* Corresponding author.

E-mail address: ratz@post.bgu.ac.il (R. Rac-Lubashevsky).

Badre (2015), for review) suggest that the control over the content of WM is achieved by an input-gating mechanism that separates the contents of WM from other sources of information, external (e.g., perceptual input) or internal (e.g., long-term memory; LTM). According to the model, when the gate to WM is closed, it provides stability by shielding the maintained information from interference caused by irrelevant or outdated information. In contrast, opening of the gate permits flexible updating of the content of WM with the new relevant input. Behavioral evidence shows that switching between these two WM functions, updating and maintenance, is associated with a cost, which reflects the time required to change the state of the gate, namely to open or close it (Kessler and Oberauer, 2014, 2015; Rac-Lubashevsky and Kessler, 2016).

One of the major challenges gating models are facing is the Homunculus problem: how is the gating mechanism controlled? According to the *prefrontal-cortex basal-ganglia working memory model* (PBWM; Frank et al., 2001; O'Reilly, 2006; Hazy et al., 2006), the basal ganglia (BG) influences WM through the dopamine (DA) circuit which projects to the prefrontal cortex (PFC). Neuropsychological data indicates that the PFC is responsible for the robust maintenance within WM (e.g., Cohen et al., 1997; Miller and Cohen, 2001; Miller et al., 1996) while the BG controls the gate (for reviews see Ranganath and Jacob (2015) and van Schouwenburg et al. (2010)). Phasic DA activity from the BG is the gating signal which opens the gate and leads to the updating of WM by disinhibiting the input from the thalamus to the PFC (Hazy et al., 2006). The gating decision is made by the same DA system through reinforcement learning which is based on the predictability and the rewarding value of the information. Specifically, an information is considered relevant if it predicts a higher reward than that predicted by the already maintained information. In such case the input will trigger gate opening. Thus the Homunculus problem is resolved by the PBWM model by assigning the DA system two roles: triggering the gating signal which opens the gate to WM, and training the DA system to predict which information would lead to future reward and should be gated into WM. This solution is in contrast to the notion of a supervisor in the form of a “central executive” (Baddeley and Hitch, 1974) that biases the contents of WM from the outside in a “top-down” manner. Instead gating models suggest that control can also be achieved in a “bottom-up” manner, through information already stored in WM and through learning mechanism in the gating system itself (Braver and Cohen, 2000; Hazy et al., 2006; O'Reilly and Frank, 2006).

The PBWM model accounts for the separation of WM representation from perceptual input by postulating a gate between two sets of representations: WM on the one hand (“higher-level”, “PFC”), and perceptually-based input (“lower-level”, “posterior”) on the other. We have recently extended this view (Rac-Lubashevsky and Kessler, 2016) by suggesting that perceptual input is not merely transient and serves for perception. Rather, perceptual processing leads to rapid and obligatory encoding into long term memory (LTM). In this case, the gate is not placed between WM and perception, but between WM and perceptually-based LTM. This extension is important because it changes the role of the closed gate. While the original gating model only regards new perceptual information as input, we suggest that recently relevant information that is now held in LTM, or information which was automatically encoded into LTM, both can be potential distractors that need to be kept outside WM. Gate closing achieves this goal. Note that gate closing is not suggested to serve also as the removal mechanism by which outdated information is removed from WM, in the same manner as gate opening is not suggested to function also as the updating mechanism of WM. Instead the two states of the gate are only assigned with the protection of the content of WM when the gate is closed and on the other hand by promoting

dynamic updating when the gate is open.

The proposal that the gate separates perceptually-based LTM from WM draws from three theoretical ideas. First, it builds on the basic tenet of multiple trace theory, which postulates that attended information is automatically encoded as episodic memories (Nadel and Moscovitch, 1997; see Moscovitch (2008) for an elaboration on the hippocampus as a “stupid” memory module). Second, it relies on the levels of processing framework (Craik and Lockhart, 1972), which regards encoding as a by-product of perceptual processing. The third is the idea that a single memory store, namely LTM, can account for all aspects of encoding and retrieval of information, without having to postulate an additional, short-term memory (STM) store (Crowder, 1982).

The idea that gating takes place between WM and LTM, rather than between WM and perception, enabled us to distinguish between LTM and WM based on the nature of updating processes that take place in these systems (Rac-Lubashevsky and Kessler, 2016). Specifically, we used the term “automatic updating” to refer to updating LTM representations. This process only depends on early selection (e.g., attending visual/auditory input), but not on late selection, which takes place based on the contents or task-relevance of the information. We argue that this is the type of memory encoding that gives rise to the recency effect, both over short or long delays. In contrast to LTM, WM updating is selective, goal-directed, depends on the content of the input (late selection), and is controlled by a gate. Moreover, the idea that LTM is constantly updated, while WM updating occurs less frequently, implies that the two systems are composed of parallel sets of representations. This is demonstrated, for example, by the fact that the recency effect, which relies on LTM, is resistant to WM load (Baddeley and Hitch, 1974). Importantly, the two sets of representations affect performance in parallel. This is most clearly observed in situations where the representations in LTM and WM are incompatible, such as in the case of n-back intrusions (see below).

1.1. The n-back paradigm

The n-back paradigm is widely used for investigation of WM updating (e.g., Jonides et al., 1997; Kirchner, 1958; Pelegriana et al., 2015). When performing the n-back task, participants are required to decide in each trial whether the presented item is same or different from the one presented *n* trials before. Typically, *n* is varied between 1 and 3 to manipulate both the demand for storage and the complexity of the updating operations, frequently referred to as “processing load”, an ambiguous and non-specific term. This task is popular in studies examining variations in WM functioning associated with aging (e.g., Oberauer, 2005; Schmie-dek et al., 2009), psychopathology (e.g., Harvey et al., 2005; Segal et al., 2015; Schoofs et al., 2008) and intelligence (Pereg et al., 2013; Price et al., 2014), and is frequently utilized in cognitive training (e.g., Dahlin, et al., 2008; Jaeggi, et al., 2010). Yet, very little is known about what exactly is measured in the n-back task.

Successful performance in n-back is supported by a complex set of cognitive processes which are hard to untangle (Jonides et al., 1997; Chatham et al., 2011) and rely on numerous brain regions (Chen et al., 2008; Jansma et al., 2000; Owen et al., 2005). These processes include: encoding the new item to WM; binding the item to its position within the set of items held in WM; comparing the new item to the one that appeared *n* trials before; inhibiting irrelevant distraction to the comparison process which stems from other items inside or outside WM (Szmalc et al., 2011); updating the item-position associations of all items in WM items (e.g., the item in position *n* – 1 should be associated with position *n* – 2 when a new item enters WM); and removing outdated information from positions that are no-longer relevant (i.e., items that

appeared more than n trials ago). The variety of processes involved in the task makes it difficult to disentangle the contribution of each of them and understand their role in overall n -back performance, as well as in individual differences in this task. In addition, it is still unknown which of these processes operate in parallel and which take place serially. Finally, pinning down the specific processes involved in n -back and their relationship to task performance is critical for understanding which processes are affected, and in what way, by manipulation of the ' n ', and as importantly—which processes are unaffected.

An important aspect of n -back performance is the need to compare each item to the one stored in the relevant position in WM (namely, the n -back item), and to resist interference from items stored in other locations. One source of such an interference are items in positions that fall beyond the minimal memory set needed to perform the task (e.g., the 3-back item in a 2-back task). These items are commonly referred to as $n+$ lures (e.g., Szmalec et al., 2011), which generate proactive interference. When these lures match the presented probe, they give rise to a tendency to respond "same", due to familiarity. Since this response tendency might be in conflict with the correct response, it must be overcome by recollection. In this context, familiarity is an automatic process which is based on activated representation in LTM while recollection is based on the bindings between an item and its position in WM (Oberauer, 2005; Szmalec, et al., 2011).

The aim of the present study was to fractionate the n -back task to its process components and to examine the contribution of each of these to individual differences in the task. To this end, we used a newly developed version of the n -back – the *reference-back* paradigm (Rac-Lubashevsky and Kessler, 2016).

1.2. The reference-back paradigm

The reference-back is composed of two types of trials: *comparison trials*, which only require a matching decision, and *reference trials*, which require both matching and updating. In each trial, participants are either presented with the stimulus "X" or "O". In contrast to the classic n -back task, the stimulus appears inside either a red or a blue frame. The participants are required to indicate whether the presented stimulus is the same as, or different from, the previous stimulus that was presented inside a red frame. In other words, only stimuli that are presented inside a red frame serve as references for comparison in future trials. For this reason, the participants should update their WM with those stimuli. These are termed *reference trials*. By contrast, stimuli appearing inside a blue frame should be compared to the previous red-framed reference, but should not be updated into WM. Those are termed *comparison trials*.

For example, assume that X was presented inside a red frame in the previous trial. Then, O appears inside the blue frame. The participant is required to respond "different", indicating that the present stimulus ("O") is different from the reference ("X"). However, since O appears inside the blue frame, it should not be updated into WM, and the reference for future trials remains X. Then, suppose that O is presented again inside a blue frame. The participant should respond again "different", indicating that the present stimulus is different from the reference (the previous item that appeared inside the red frame, namely "X") even though it matches the previous immediate trial, which was also an O. Then, suppose that O is presented inside a red frame. The participant should respond "different", however, this time, since O was presented inside the red frame, WM updating is required, and the reference, to which the following trials should be compared, becomes O (see Fig. 1).

In terms of face validity, the reference-back task is similar to the classic n -back in the sense that they both require frequent updating of WM contents, as well as categorizing each stimulus as same/different in comparison to previous trials. However, the reference-back task solves an inherent problem that exists in n -back. Specifically, it is impossible to measure and manipulate the updating process in isolation: since each trial in the n -back task requires updating of WM with the presented stimulus, the task lacks a baseline for comparison. The reference-back task solves this problem by including trials in which a comparison to previous trials in WM is to be carried out, without requiring the process of WM updating. By doing so, it enables the identification of the contribution of additional processes, such as gating and the comparison process ("matching").

1.3. Sub-processes derived from the reference-back task

The design of the reference-back task enables the separation of several sub-processes that take place during task performance. We describe in the following paragraph each of these processes, their calculation, and how each relates to performance in the classic n -back task. Each sub-process will be later used as individual-differences predictor. For this reason, they were defined as orthogonal contrasts which will allow us to assess their unique role in explaining individual differences in the standard n -back task. Table 1 presents the components and their calculation in the task.

(a) Updating cost

In n -back, successful performance requires that in each trial the presented item would be updated to WM, along with its position within the WM set. Thus, updating should take place in all trials. Moreover, it involves all items, since each trial

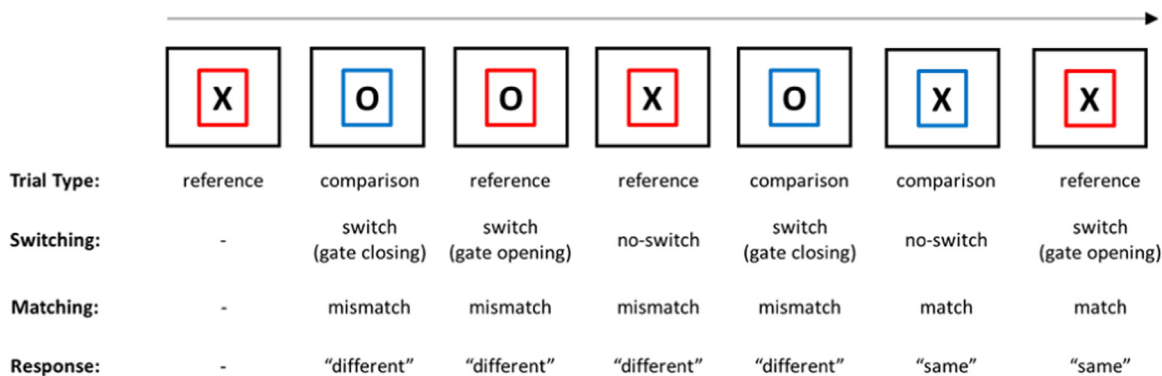


Fig. 1. The reference-back task is described. Trials with red frame are reference trials and trials with blue frame are comparison trials (see text for details). The sequence length for each trial-type was random. The titles inside the quotation marks represent the accurate response in that trial. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1

The calculation of specific predictors as orthogonal contrasts derived from the reference-back task. “Switch” and “no-switch” refer to the alternation or repetition of the trial-type, namely comparison vs. reference.

	Comparison trials				Reference trials			
	Switch		no-switch		Switch		no-switch	
	Mismatch	Match	Mismatch	Match	Mismatch	Match	Mismatch	Match
Matching			+	–			+	–
Updating cost	–	–	–	–	+	+	+	+
Substitution			–	+			+	–
Gate opening					+	+	–	–
Gate closing	+	+	–	–				

entails a change in the set of associations between items and their positions due to “pushing back” the memory set (i.e., position $n-1$ becomes $n-2$, $n-2$ becomes $n-3$ and so forth). In the reference-back task, updating only takes place in reference trials, and not in comparison trials.

It is important to note that updating is not limited to the mismatch condition, but takes place in the match condition as well. It could be argued that the 1-back is an exception to the “pushing back” rule, because updating is (presumably) only required in mismatch trials (e.g., X following O), but not in match trials (X following X), where the memory set is unaffected. However, previous results from the reference-back task, in which the updating condition is similar in essence to 1-back, showed that reference trials are always slower than comparison trials, irrespective of the matching condition (Rac-Lubashevsky and Kessler, 2016). These findings confirm that updating is involved in all reference trials.

The idea that updating is not limited to the mismatch condition implies that it does not only occur in situations where new information is presented. Rather, updating can take place even when the information that is entered to WM is identical to the information that is already maintained. In this case, the duration of updating would be shorter than in mismatch situations (Kessler et al., 2016; Oberauer et al., 2013).

Accordingly, the duration of updating (“updating cost”) was measured as the difference in performance between reference trials and comparison trials (regardless of matching conditions) which served as the baseline for updating.

(b) *Substitution*

While updating cost reflects the duration of updating *any* item in WM, substitution reflects the additional time required to update WM with new information. As explained above, updating only takes place in reference trials. Whereas updating cost was calculated beyond the match/mismatch condition, substitution was calculated as the difference between the matching component in reference trials and the matching component in comparison trials. Specifically, it is calculated as

$$\begin{aligned}
 & (\text{Mismatch}_{\text{reference}} - \text{Match}_{\text{reference}}) \\
 & - (\text{Mismatch}_{\text{comparison}} - \text{Match}_{\text{comparison}}) = \text{Mismatch}_{\text{reference}} \\
 & \quad - \text{Match}_{\text{reference}} \\
 & \quad - \text{Mismatch}_{\text{comparison}} + \text{Match}_{\text{comparison}}.
 \end{aligned}$$

Ecker et al. (2010) decomposed WM updating to three components: substitution, retrieval and transformation. Our definition of substitution is the same as Ecker et al.’s, being the replacement of the content of memory with new information.

(c) *Matching (also termed “comparison”)*

This is the process by which the presented stimulus is compared to the relevant reference item in WM, resulting in a

match (a “same” response) or a mismatch (“different”) decision (Chen et al., 2008). In the standard n-back task, matching should always refer to the stimulus that appeared n trials before. In the reference-back task, matching should refer to the most recent reference trial, namely the previous stimulus that appeared inside a red frame. Note that this feature makes the reference-back similar to a 1-back task, since they both require maintaining only one item in WM. In 1-back this is always the item that appeared in the immediately previous trial, while in the reference-back task it could be encoded to WM several trials before, i.e., when the last red framed item appeared.

In order to provide a baseline for substitution, we will examine the additional time required to achieve a mismatch decision (“different”) compared to match. This will be calculated beyond the trial type (so it will be orthogonal to all other predictors; see Table 1), namely in both reference and comparison trials. It should be noted that longer RTs for mismatch decisions is a well-grounded finding in the literature, which is not limited to tasks that require WM updating (for early treatments see Eriksen et al. (1982), Farrell (1985) and Sternberg (1966)). In addition, the matching process seems to be independent of ‘ n ’: Watter et al. (2001) demonstrated that the peak P300 amplitude in the n-back task was larger in match compared to mismatch trials, but its latency was unaffected by ‘ n ’ manipulation (1-, 2-, or 3-back).

(d) *Gate opening*

As suggested above, in order for actual updating of information to take place, prior opening of the gate to WM is required. Previous work by Kessler and colleagues has shown that switching between updating and maintenance, or vice versa, is associated with an RT cost (Kessler and Oberauer, 2014, 2015; Rac-Lubashevsky and Kessler, 2016). In the present work, the duration of gate opening was measured as the performance difference between reference trials that were preceded by comparison trials (“switch trials”), and reference trials preceded by reference trials (“no-switch trials”). Whereas both situations require keeping the gate open, only switch trials involve the additional process of opening the gate. This is consistent with the findings of Kessler and Oberauer (2014, 2015), showing that the present “state” of the gate does not change along successive trials, unless such a change is required.

(e) *Gate closing*

Gate closing is the opposite effect of gate opening. The PBWM model holds that the gate is closed by default, and hence transient opening should always be accompanied by an autonomous gate closing. However, this is not necessarily the case, especially in situations where both updating and maintenance are frequently and equally required. For example, in a letter updating task, Kessler and Oberauer (2014, 2015)

showed that both gate opening and closing are costly in terms of RT. Moreover, Rac-Lubashevsky and Kessler (2016) showed a larger cost for gate closing than for opening.

Empirically, gate closing is calculated as the difference in performance between comparison trials that were preceded by reference trials, ("switch trials") and comparison trials that are preceded by comparison trials ("no-switch trials"). We suggest that in switch trials the gate closes in order to protect the reference stimulus from being overwritten by the information presented in the following comparison trials. In the subsequent no-switch trials the gate remains closed.

(f) Intrusion

While gate closing deals with interference that stems from outside the WM set ($n+$ lures), the gate is not predicted to control interference generated within WM, such as from $n-$ lure trials. This interference effect, also termed intrusion (Oberauer, 2005), is generated from lure trials which are part of the maintained memory set, but are still not the relevant reference for comparison (e.g., the 1-back item in a 2-back task). There is a possibility that $n-$ lures are controlled by an additional output gate (Chatham et al., 2014), however more evidence are needed to support the existence of such a mechanism. Furthermore, intrusion effect is substantially larger in magnitude than interference from $n+$ trials (Chatham et al., 2011; Szmalec et al., 2011). In the present study, intrusion is measured in a 2-back task by comparing mismatch decisions in which the 1-back item is a lure (e.g., XOO), to mismatch decisions without interference from trial $n-1$ (e.g., XXO).

1.4. The present study

The aim of the present study is to examine the unique contribution of the mentioned processes to performance in the task. To this end, participants were tested in the reference-back task, which allowed extraction of estimates for matching, substitution and gating. These served to explain individual differences in performance on the standard 1-back and 2-back tasks. Intrusion was added as a predictor for explanation of variance in the 2-back task. In an additional analysis, the predictors above were used in order to understand the difference(s) in performance between 2-back and 1-back, aiming to clarify which processes are specifically manipulated or added when increasing the n . To account for overall processing speed, all the individual differences models were tested against performance in a 2-choice RT task, in which participants were required to classify the stimulus identity (e.g., $X \rightarrow$ left response, $O \rightarrow$ right response), with minimal involvement of WM.

2. Method

2.1. Participants

90 undergraduate students from Ben-Gurion University of the Negev participated in the experiment (32 males; age: $M = 24.66$, $S.D. = 2.75$). Informed consent was obtained from all participants and they were paid for their participation. Two participants were removed from the analysis, one due to extremely low accuracy in the reference back task ($< 50\%$ in some of the conditions), and the other due to extreme RT outliers.

2.2. Stimuli and apparatus

Stimuli presentation and behavioral data collection were done using E-Prime v2.0 (Psychology Software Tools, Pittsburgh, PA).

The stimuli were the letters "X" and "O", in font size 36, presented in black against a light gray background within a red or a blue frame. Responses were collected using a response box.

2.3. Procedure

The experiment was composed of two sessions, carried out in different days, not more than a week apart. In the first session, participants performed the 2-choice RT task, followed by the 1-back and 2-back tasks. Then, they performed the practice phase of the reference-back task. In the second session, participants performed only the reference-back task.

In all tasks, each trial started with a fixation screen that was presented for 800 ms, followed by a blank display for 1000 ms. Then, one of the stimuli "X" or "O" was selected at random and presented until a response was indicated. Participants were instructed to keep their eyes fixated on the center of the screen throughout the experimental blocks. In the 2-choice RT task, participants were asked to indicate whether the stimulus was an "X" or an "O" (if X press left, if O press right). In the 1-back task participants were required to decide whether each stimulus was the same or different from the stimulus presented in the previous trial. In the 2-back task, participants were required to decide whether each stimulus was the same or different from the stimulus presented two trials beforehand. The reference-back instructions are described below. "Same" and "different" responses were indicated by using the right and left index fingers respectively, on a response box. The 2-choice RT task was composed of 2 blocks including 60 trials each, preceded by a 30-trial practice block. The 1-back task was composed of a practice block followed by 3 experimental blocks, including 60 trials each. The 2-back task was composed of one practice block and 4 experimental blocks, including 60 trials each.

2.3.1. The reference-back task

The reference-back task was composed of two trial types: *reference trials*, indicated by a red frame around the stimulus, and *comparison trials*, indicated by a blue frame (see Fig. 1). The stimulus in each trial was an "X" or an "O", chosen at random. In 75% of the trials the trial type (reference or comparison) was the same as in the preceding trial. In the remaining 25% of the trials, the trial type was alternated.

In each trial, participants had to indicate whether the stimulus is the same or not as the one that appeared in the *most recent red frame* (i.e., the last reference trial). "Same" and "different" responses were indicated using the right and left index fingers respectively. Each block contained 60 trials. Participants completed four practice blocks as part of the first session, and 15 experimental blocks as part of the second session.

3. Results

All the analyses were carried out using Bayesian statistics, and implemented using the JASP (Love et al., 2015) statistical software. One advantage of Bayesian statistics over null hypothesis significance testing is in its ability to gather evidence for both the null and the alternative hypotheses. We based our decisions on the Bayes Factor (BF) statistic, which is the ratio between the probability of the data given the alternative hypothesis (H_1) and the probability of the data under the null hypothesis (H_0). For example, $BF = 20$ indicates that the data is twenty times more likely under the alternative hypothesis, assuming a difference between the conditions, than it is under the null hypothesis. Unlike the p -value, the interpretation of the BF as a measure of evidence is straightforward. Moreover, instead of making a binary decision,

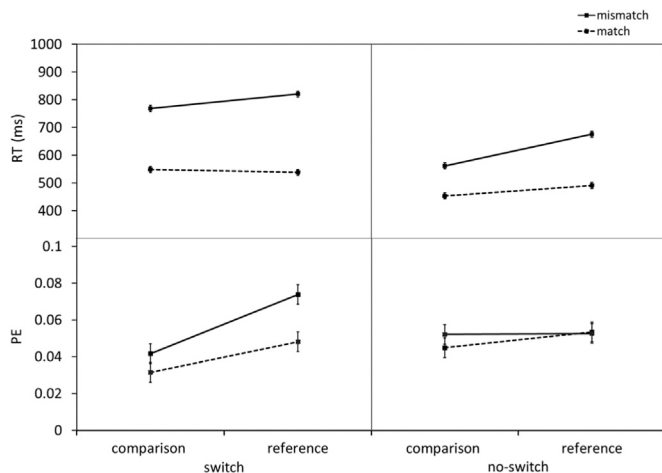


Fig. 2. The 3-way ANOVA between Trial-type (comparison/reference), Gate-Switch (switch, no-switch) and Matching (mismatch, match) is presented for reaction time (RT) and error proportion (PE). RT difference between mismatch and match trials is larger in reference trials than in comparison trials, reflecting the additional time taken for substitution in this condition. A larger gate switching cost (difference between switch and no-switch) is observed in comparison trials (corresponding to gate closing) than in reference trials (gate opening). Error bars represent 95% confidence intervals (Masson and Loftus, 2003).

BFs provide a quantitative estimate of the evidence. According to commonly accepted norms (Jeffreys, 1961), BFs between 1 and 3 are regarded as merely anecdotal evidence for H_1 , between 3 and 10 as substantial, between 10 and 30 as strong, between 30 and 100 as very strong, and BFs larger than 100 as decisive.

3.1. Reference-back performance analysis

RT and PE in the reference-back task are presented in Fig. 2 as a function of Trial-Type (comparison, reference), Gate-Switch (switch, no-switch) and Matching (mismatch, match). We focused only on the specific components described above. A full exploration of the group-level results is not the focus of the present paper, and is available elsewhere (Rac-Lubashevsky and Kessler, 2016). The presents of the relevant sub-processes of the reference-back were tested individually to base their later use as predictors of the performance in n-back. One-sample Bayes Factor (BF) for each component, was computed to ensure that the mean of each component is reliably different than zero. Table 2 for reaction time (RT) and Table 6 for error proportion (PE).

3.2. Individual differences analysis

3.2.1. RT

The RT data was log-transformed in order to normalize the data (see Table 2 for skewness and kurtosis). Predictors that were

calculated as difference scores were calculated as the log of the ratio between the two conditions (e.g., mismatch cost was calculated as $\log_{10}(\text{mismatch}/\text{match})$). Bivariate correlations between all study variables in RT are presented in Table 3.

3.2.2. Bayesian Linear Regression

Next, Bayesian Linear Regression analyses were conducted for predicting 1-back and 2-back RT, as well as the RT difference between 1- and 2-back, from the process components. All substantial models, which include a subset of the predictors derived from the reference-back task, were compared to a null model which included only mean RT in 2-choice RT in order to control for overall individual differences in processing speed.

The model selection procedure was as follows. We started by selecting the best fit model based on the highest BF among all possible models. It should be noted, however, that the BF ratio between this model and the second-best model was not always substantial (i.e., not larger than 3). In these cases, we preferred the more parsimonious model, and hence selected the second-best model if it included fewer predictors. Individual differences in 1-back were explained by the duration of updating (updating cost) and gate opening, $BF=29,972.54$ ($R^2=.53$). RT in 2-back was explained by updating, gate closing and intrusion, $BF=209.00$ ($R^2=.37$) while the RT difference between the 1- and 2-back was explained only by gate closing and intrusion, $BF=10.86$ ($R^2=.16$). The standardized regression coefficients of the selected predictors are presented in Table 4. Table 5 presents the posterior probability and BF for including each predictor, summed across all tested models. Model averaging is a complementary and alternative approach to model selection, in which the posterior probability of each predictor, rather than of each model, is summed across all possible models. BF is then calculated as the posterior to prior odds ratio (Rouder et al., 2016).

3.2.3. PE analysis

To determine which of the calculated components predicts successful performance in 1-back and 2-back, a regression analysis was done on PE. The PE data was transformed using arc sin transformation in order to normalize the data (see Table 6 for descriptive statistic of PE data). Bivariate correlations between all study variables in PE are presented in Table 7.

3.2.4. Bayesian Linear Regression

As in the RT analysis, the substantial models were tested against a null model with PE in the 2-choice task, which only accounts for overall individual differences in error (see Tables 4 and 5).

Individual differences in 1-back were best explained by the matching and gate opening predictors, $BF=4.86$ ($R^2=.29$) while error in 2-back, and the error in the difference between 1-back and 2-back, were explained by the intrusion predictor alone,

Table 2

Descriptive statistics for the RT data. Skewness, kurtosis and reliability were calculated after log transformation. Split-half reliability was calculated as the correlation between the first and second half of the experiment, and corrected using the Spearman-Brown formula. BF=one-sample Bayes Factor against zero.

	Mean (ms)	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Reliability	BF
Matching	138	87	14	418	.57	-.17	.95	2.13×10^{22}
Updating cost	62	53	-28	240	.05	-.07	.85	2.01×10^{15}
Substitution	74	99	-109	429	.40	.17	.86	2.39×10^7
Gate opening	93	114	-42	529	.78	.45	.86	3.68×10^8
Gate closing	143	105	8	560	.50	.36	.86	3.98×10^{18}
Intrusion	374	425	-119	1641	.09	-.26	.67	5.39×10^9
1-back	554	141	318	1125	.08	-.14	.99	
2-back	1141	612	399	3044	.20	-.82	.99	
2-back minus 1-back	587	524	-84	2336	.45	-.18	.98	
2-choice RT	380	59	287	576	.44	.01	.99	

Table 3
Correlation table between the log-transformed RT predictors. BF were calculated compared to the null hypothesis that assumes no correlation, and are reported in parentheses. Correlations with BF larger than 3 are boldfaced.

Bayesian Pearson Correlations for RT										
		Updating cost	Substitution	Gate opening	Gate closing	Intrusion	1-back	2-back	2-back minus 1-back	2-choice RT
Matching	Pearson's r	.43	.24	.30	.16	–.12	.33	.26	.14	.41
	BF	(709.40)	(1.63)	(6.72)	(.38)	(.24)	(18.18)	(2.45)	(.30)	(362.17)
Updating cost	Pearson's r		.30	.33	.30	–.21	.41	.36	.23	.21
	BF		(6.65)	(19.34)	(7.67)	(.92)	(287.28)	(50.53)	(1.18)	(.90)
Substitution	Pearson's r			.25	–.05	.06	.27	.22	.12	.15
	BF			(1.90)	(.15)	(.16)	(3.46)	(1.02)	(.25)	(.34)
Gate opening	Pearson's r				.43	–.14	.54	.40	.19	.26
	BF				(852.27)	(.30)	(200,648.44)	(199.05)	(.60)	(2.47)
Gate closing	Pearson's r					–.32	.41	.36	.22	.16
	BF					(10.80)	(320.98)	(47.79)	(1.00)	(.39)
Intrusion	Pearson's r						–.14	.08	.20	–.03
	BF						(.33)	(.17)	(.81)	(0.14)
1-back	Pearson's r							.72	.33	.58
	BF							(5.31 × 10¹²)	(16.81)	(4.54 × 10⁶)
2-back	Pearson's r								.89	.44
	BF								(4.47 × 10²⁷)	(991.66)
2-back minus 1-back	Pearson's r									.22
	BF									(.10)

Table 4
Standardized regression coefficients are presented for each predictor in the selected model predicting RT and PE in 1-back, 2-back and difference between 2-back and 1-back.

Predictor	β		
	1-back	2-back	2-back minus 1-back
RT			
2-choice RT	.45	.34	.18
Updating cost	.20	.25	
Gate opening	.35		
Gate closing		.31	.28
Intrusion		.24	.30
PE			
2-choice RT	.43	.32	.15
Matching	.20		
Updating cost	–.19		
Intrusion	–.15	.21	.26

BF=1.84 ($R^2=.15$) and BF=4.21 ($R^2=.09$), respectively. Overall, BFs for the PE regression models are very small, especially compared to those of the RT models. This could be partially explained

by the fact that errors in this task only reflect a faulty matching decision, either generally (as in 1-back) or due to referring to a wrong reference (as in intrusion trials). However, errors in updating or gating in a given trial, such as failing to update in a reference trial or an erroneous updating in a reference trial, do not affect performance in that trial but only in the following trials.

4. General discussion

The aim of the current study was to examine the contribution of matching, updating, substitution, gate opening and gate closing to variance in performance on the n-back task, as well as the contribution of intrusion specifically to 2-back performance.

Chatham et al. (2011) have proposed a biological model that explains individual difference in n-back performance, based on the PBWM model. They demonstrated that the strategy used by the participants, namely relying on familiarity (i.e., LTM) or on recollection (i.e., WM) in order to remember the n-back stimulus, explained their ability to perform well on trials that required resistance to proactive interference. Variations in the ability to resist

Table 5
The prior and posterior probabilities of each predictor, in the Bayesian regression analysis, summed across all tested models for RT and PE in 1-back, 2-back and difference between 2-back and 1-back. P (incl) represents the prior probability for each predictor. This probability is equal for all the predictors. P (incl|data) represents the posterior probability of each predictor given the data. BF_{inclusion} is calculated as the odds ratio between the posterior and prior probabilities. BFs larger than 3 are boldfaced.

Model	P (incl)	1-back		2-back		2-back minus 1-back	
		P (incl data)	BF _{Inclusion}	P (incl data)	BF _{Inclusion}	P (incl data)	BF _{Inclusion}
RT							
(1) Matching	.5	.18	.23	.23	.31	.28	.40
(2) Updating cost	.5	.49	1.47	.71	2.43	.57	1.33
(3) Substitution	.5	.36	.56	.35	.54	.31	.44
(4) Gate opening	.5	.98	45.45	.57	1.35	.32	.48
(5) Gate closing	.5	.65	1.83	.84	5.31	.72	2.62
(6) Intrusion	.5	/	/	.73	2.76	.89	7.94
PE							
(1) Matching	.5	.75	3.02	.31	.44	.51	1.03
(2) Updating cost	.5	.33	.49	.42	.72	.38	.60
(3) Substitution	.5	.25	.33	.38	.62	.38	.61
(4) Gate opening	.5	.51	1.03	.28	.38	.29	.42
(5) Gate closing	.5	.40	.68	.34	.52	.31	.45
(6) Intrusion	.5			.64	1.80	.77	3.40

Table 6

Descriptive statistics for the PE data. Skewness, kurtosis and reliability were calculated after arc sinus transformation. Reliability was calculated as the correlation between the first and second half of the experiment, and corrected using the Spearman-Brown formula. BF = one-sample Bayes Factor against zero.

	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Reliability	BF
Matching	0	.03	−.10	.05	−.71	1.65	.57	.44
Updating cost	.01	.02	−.06	.05	−.82	1.81	.72	148,002.22
Substitution	0	.05	−.13	.21	.79	3.10	.50	.12
Gate opening	0	.03	−.06	.14	1.25	3.17	.61	.14
Gate closing	.01	.03	−.11	.05	−1.08	2.62	.48	1.21
Intrusion	.02	.09	−.13	.30	1.04	1.09	.69	1.06
1-back	.05	.03	.01	.20	1.59	3.79	.82	
2-back	.12	.09	.03	.41	1.30	1.15	.89	
2-back minus 1-back	.07	.08	−.06	.37	1.12	1.36	.82	
2-choice RT	.01	.02	.00	.08	1.84	4.45	.66	

interference was explained by genetic variations of tonic DA level. Distractor-resistant (robust) maintenance in the PFC is supported by high tonic DA levels, which foster stability over flexibility, rendering WM updating more difficult. In contrast, low tonic DA levels can be easily overridden by phasic DA bursts, which leads to a weaker ability to provide stability and resist interference and thus supports rapid and (over) flexible updating. While these results point to an important source of individual difference in n-back they still miss other possible sources of individual differences in the task which were addressed in the present study.

Our RT results demonstrated that updating contributed in explaining individual differences in 1-back and 2-back. At the very basic level, these results provide empirical validation for the n-back as a WM updating paradigm. This is noteworthy since updating cost, which reflects the additional time required to update WM compared to LTM, contributes to individual difference even in 1-back, which is sometimes referred to as a task that predominantly requires maintenance, with no or little need for updating (e.g., Ragland et al., 2002). Moreover, not only that updating is required in 1-back, control is required as well in order to enable the updating state. The presence of control even in 1-back is supported by the correlation and regression results showing that gate opening together with updating explains a good part (53%) of the variance in 1-back.

Furthermore, the fact that gate opening was included only in the 1-back model but not in the 2-back models can be explained by individual differences in strategy, which may occur on 1-back

but not on 2-back: In 1-back, updating is mandatory only in mismatch trials, while in 2-back it is mandatory in every trial. As was mentioned before, updating was shown to occur even when substitution was not needed (Rac-Lubashevsky and Kessler, 2016). However, it is possible that participants might differ in their decision to update (as detected by gate opening) in matches in 1-back, while in 2-back no difference in updating strategy should take place.

RT in 2-back was explained by updating, gate closing and intrusion while the RT difference between the 1- and 2-back was explained only by gate closing and intrusion. Although the bivariate correlation between intrusion and 2-back RT was neglectful, intrusion *did* explain part of the variance in 2-back when *gate closing was controlled for*. This indicates that a suppression situation might have occurred (e.g., Tzelgov and Henik, 1991), i.e., that gate-closing suppressed the relation between intrusion and 2-back. While intrusion is negatively correlated with gate closing, gate closing is positively correlated with 2-back. Thus, only when considering the unique contribution of intrusion (“clean” from the shared variance with gate closing), the positive correlation with 2-back was detectable. This suppression could be explained by the opposite effect that gate closing and intrusion could have on processing time in 2-back. Large intrusion cost will increase RT while small or no gate closing time will benefit RT. This pattern could be the results of participants who are low in their ability to resist proactive interference. In this case these participant would have less ability to deal with intrusion (and will have larger

Table 7

Bayesian Pearson Correlations table between predictors calculated from PE after arc sinus transformation. BF were calculated compared to the null hypothesis that assumes no correlation, and are reported in parentheses. Correlations with BF larger than 3 are boldfaced.

Bayesian Pearson Correlations for PE									
		Updating cost	Substitution	Gate opening	Gate closing	Intrusion	1-back	2-back	2-back minus 1-back
Matching	Pearson's r	.17	−.06	−.09	−.12	−.06	−.34	.50	.14
	BF	(.46)	(.15)	(.19)	(.26)	(.15)	(21.73)	(27,945.00)	(.32)
Updating cost	Pearson's r		−.29	−.35	.49	−.09	−.16	−.15	−.11
	BF		(4.68)	(36.31)	(12,478.24)	(.19)	(.38)	(.34)	(.22)
Substitution	Pearson's r			.42	−.11	−.02	.13	.15	.12
	BF			(466.28)	(.23)	(.13)	(.27)	(.37)	(.25)
Gate opening	Pearson's r				−.41	.08	.22	.14	.06
	BF				(265.72)	(.18)	(1.14)	(.29)	(.15)
Gate closing	Pearson's r					−.04	−.13	−.12	−.08
	BF					(.14)	(.29)	(.25)	(.18)
Intrusion	Pearson's r						.02	.23	.27
	BF						(.13)	(1.35)	(3.03)
1-back	Pearson's r							.50	.14
	BF							(27,945.00)	(.32)
2-back	Pearson's r								.93
	BF								(5.18 × 10³⁴)
2-back minus 1-back	Pearson's r								.17
	BF								(.43)

intrusion cost) and they will also be less likely to close the gate (leading to a smaller closing cost overall). These results suggest then that the ability to resist irrelevant information from within WM is related to the ability to resist irrelevant perceptual-LTM based information. However, this interpretation still needs to be confirmed by additional data.

The presence of control in n-back was demonstrated by our results as gating, being the control mechanisms over WM content, was a strong predictor in all the selected models. Thus, n-back can be described as paradigm that involves conflict (Oberauer, 2005). In fact, the level of the conflict can be manipulated by the size of the stimulus set used in the task. In small stimulus sets, match-to-target trials are more frequent, making it more difficult to recollect the precise position of recent trials (Gray et al., 2003; Heathcote et al., 2015; Szmalec et al., 2011). Gate closing is proposed here to be the mechanism that can reduced the level of conflict created in these situations.

It is worthwhile to note that, substitution, defined as the additional time taken to update WM with a new information compared to updating it with an already-existing information, was not included in any of our best-fitting models. This finding that individual differences in substitution do not play a major role in overall n-back performance is consistent with the results of Ecker et al. (2010), where substitution was not correlated with WM capacity. However, because n-back is not a “WM capacity task” but is rather generally conceived as an updating task, our results are somewhat surprising. Our results indicate that substitution contributes very little to the overall individual differences in n-back. It is possible that substitution is a relatively “low-level” process and updating is better defined as a state of the WM system, captured by the difference between being in an updating (reference trials) or maintenance (comparison) state, then by the specific incoming stimulus.

The correlation results demonstrated a positive correlation between gate opening and closing in RT. This correlation might be attributed to a conflict management time component. Switch trials, in both reference and comparison trials, demand a decision on whether WM updating is relevant or not. The positive correlation shows that changing the state of the gate between opened and closed, is symmetrical in the time that it takes. However, the association between gate opening and closing was negative in PE. It is possible that this negative correlation is the result of a personal tendency or strategy favoring either stability or flexibility (Braver, 2012; Hommel, 2015). According to this view, a strong bias towards one state of the gate might lead to more errors under conditions which require the other state. For example, a low criterion for opening the gate will lead to a dominant flexible behavior. This behavior is appropriate in circumstances that demand frequent updating, yet it could also lead to distractibility and error under circumstances which require stability. In contrast, high criterion for opening the gate will lead to dominant persistency. This behavior will lead to successful performance under conditions that require robust maintenance but it could also lead to maladaptive rigidity stemming from prevention of updating (Dreisbach and Goschke, 2004; Miller and Cohen, 2001; Meiran et al., 2011). However, the reliability of the PE results is weak, and therefore this interpretation should be taken with caution.

To conclude, WM updating has long been recognized as a core executive function (Kane and Engle, 2002; Miyake and Friedman, 2012). However, part of what makes updating an executive operation is the control that it requires. Because the control over WM is inherent when measuring updating of WM, it makes the direct measurement and manipulation of updating complicated. The present study utilized the reference-back paradigm, which enables to differentiate empirically between these two processes and also with other processes such as matching and intrusion and

examined their unique contribution to n-back, presumably the most widely-used WM updating task. Based on our decomposition of the task, we found that control, as implemented by the gating mechanism serves as the main source of individual differences in n-back, including in 1-back. Future work is required to identify the neural substrates of each process, and its relationship to other executive functions (e.g., shifting, inhibition), as well as to other sources of individual differences such as WM capacity and intelligence. Due to its high specificity, we suggest that the reference-back task can be used for characterization, and possibly even diagnosis of deficits in WM control and WM-updating, which are associated with a variety of psychological and neuropsychological conditions.

Acknowledgment

This research was funded by the Israel Science Foundation (Grant #458/14) awarded to Y. K. We thank Tal Yatziv for helpful comments on the manuscript. We also thank Avital Frenkel and Ido Geva for their assistance with data collection.

References

- Baddeley, A., 1992. Working memory. *Science* 255, 556–559.
- Baddeley, A.D., Hitch, G.J., 1974. Working memory. *Psychol. Learn. Motiv.* 8, 47–89.
- Bays, P.M., Husain, M., 2008. Dynamic shifts of limited working memory resources in human vision. *Science* 321, 851–854.
- Braver, T.S., 2012. The variable nature of cognitive control: a dual mechanisms framework. *Trends Cognit. Sci.* 16, 106–113.
- Braver, T.S., Cohen, J.D., 2000. On the control of control: the role of dopamine in regulating prefrontal function and working memory. In: Monsell, S., Driver, J. (Eds.), *Attention and Performance XVIII: Control of Cognitive Processes*. MIT Press, Cambridge, MA.
- Chatham, C.H., Badre, D., 2015. Multiple gates on working memory. *Curr. Opin. Behav. Sci.* 1, 23–31.
- Chatham, C.H., Frank, M.J., Badre, D., 2014. Corticostriatal output gating during selection from working memory. *Neuron* 81, 930–942.
- Chatham, C.H., Herd, S.A., Brant, A.M., Hazy, T.E., Miyake, A., O'Reilly, R., Friedman, N.P., 2011. From an executive network to executive control: a computational model of the n-back task. *J. Cognit. Neurosci.* 23, 3598–3619.
- Chen, Y.N., Mitra, S., Schlaghecken, F., 2008. Sub-processes of working memory in the n-back task: an investigation using ERPs. *Clin. Neurophysiol.* 119, 1546–1559.
- Cohen, J.D., Perlstein, W.M., Braver, T.S., Nystrom, L.E., Noll, D.C., Jonides, J., Smith, E. E., 1997. Temporal dynamics of brain activation during a working memory task. *Nature* 386, 604–608.
- Craik, F.I., Lockhart, R.S., 1972. Levels of processing: a framework for memory research. *J. Verbal Learn. Verbal Behav.* 11, 671–684.
- Crowder, R.G., 1982. The demise of short-term memory. *Acta Psychol.* 50, 291–323.
- Dahlin, E., Neely, A.S., Larsson, A., Bäckman, L., Nyberg, L., 2008. Transfer of learning after updating training mediated by the striatum. *Science* 320, 1510–1512.
- Dreisbach, G., Goschke, T., 2004. How positive affect modulates cognitive control: reduced perseveration at the cost of increased distractibility. *J. Exp. Psychol.: Learn. Mem. Cogn.* 30, 343–353.
- Ecker, U.K., Oberauer, K., Lewandowsky, S., 2014. Working memory updating involves item-specific removal. *J. Mem. Lang.* 74, 1–15.
- Ecker, U.K.H., Lewandowsky, S., Oberauer, K., 2014. Removal of information from working memory: a specific updating process. *J. Mem. Lang.* 74, 77–90.
- Ecker, U.K.H., Lewandowsky, S., Oberauer, K., Chee, A.E., 2010. The components of working memory updating: An experimental decomposition and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 36, 170–189.
- Eriksen, C.W., O'hara, W.P., Eriksen, B., 1982. Response competition effects in same-different judgments. *Percept. Psychophys.* 32, 261–270.
- Farell, B., 1985. “Same”–“different” judgments: a review of current controversies in perceptual comparisons. *Psychol. Bull.* 98, 419.
- Frank, M.J., Loughry, B., O'Reilly, R.C., 2001. Interactions between frontal cortex and basal ganglia in working memory: a computational model. *Cognitive, Affective, and Behavioral Neuroscience* 1, 137–160.
- Gray, J.R., Chabris, C.F., Braver, T.S., 2003. Neural mechanisms of general fluid intelligence. *Nat. Neurosci.* 6, 316–322.
- Harvey, P.O., Fossati, P., Pochon, J.B., Levy, R., LeBastard, G., Lehericy, S., Aljlilare, J.F., Dubois, B., 2005. Cognitive control and brain resources in major depression: an fMRI study using the n-back task. *Neuroimage* 26, 860–869.
- Hazy, T.E., Frank, M.J., O'Reilly, R.C., 2006. Banishing the homunculus: making working memory work. *Neuroscience* 139, 105–118.

- Heathcote, A., Coleman, J.R., Eidels, A., Watson, J.M., Houpt, J., Strayer, D.L., 2015. Working memory's workload capacity. *Mem. Cogn.*, 1–17.
- Hommel, B., 2015. Between persistence and flexibility: the Yin and Yang of action control. *Adv. Mot. Sci.* 3, 33–67.
- Jaeggi, S.M., Studer-Luethi, B., Buschkuhl, M., Su, Y.F., Jonides, J., Perrig, W.J., 2010. The relationship between n-back performance and matrix reasoning—implications for training and transfer. *Intelligence* 38 (6), 625–635.
- Jansma, J.M., Ramsey, N.F., Coppola, R., Kahn, R.S., 2000. Specific versus nonspecific brain activity in a parametric N-back task. *Neuroimage* 12, 688–697.
- Jonides, J., Schumacher, E.H., Smith, E.E., Lauber, E.J., Awh, E., Minoshima, S., Koeppe, R.A., 1997. Verbal working memory load affects regional brain activation as measured by PET. *Journal of Cognitive Neuroscience* 9, 462–475.
- Kane, M.J., Engle, R.W., 2002. The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychon. Bull. Rev.* 9, 637–671.
- Kessler, Y., Meiran, N., 2008. Two dissociable updating processes in working memory. *J. Exp. Psychol.: Learn. Mem. Cogn.* 34, 1339–1348.
- Kessler, Y., Oberauer, K., 2014. Working memory updating latency reflects the cost of switching between maintenance and updating modes of operation. *J. Exp. Psychol.: Learn. Mem. Cogn.* 40, 738–754.
- Kessler, Y., Oberauer, K., 2015. Forward scanning in verbal working memory updating. *Psychon. Bull. Rev.* 22, 1770–1776.
- Kessler, Y., Baruchin, L. J., & Bouhsira-Sabag, A. (in press). Working memory updating occurs independently of the need to maintain task-context: accounting for triggering updating in the AX-CPT paradigm. *Psychological Research*. 1–13.
- Kirchner, W.K., 1958. Age differences in short-term retention of rapidly changing information. *Journal of experimental psychology* 55, 352–358.
- Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., Ly, A., Gronau, Q. F., Smira, M., Epskamp, S., Matzke, D., Wild, A., Rouder, J. N., Morey, R.D., Wagenmakers, E.-J., 2015. JASP (Version 0.7) (Computer software).
- Luck, S.J., Vogel, E.K., 1997. The capacity of visual working memory for features and conjunctions. *Nature* 390, 279–281.
- Luck, S.J., Vogel, E.K., 2013. Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends Cognit. Sci.* 17, 391–400.
- Ma, W.J., Husain, M., Bays, P.M., 2014. Changing concepts of working memory. *Nat. Neurosci.* 17, 347–356.
- Masson, M.E.J., Loftus, G.R., 2003. Using confidence intervals for graphically based data interpretation. *Can. J. Exp. Psychol.* 57, 203–220.
- Meiran, N., Diamond, G.M., Toder, D., Nemets, B., 2011. Cognitive rigidity in unipolar depression and obsessive compulsive disorder: examination of task switching, Stroop, working memory updating and post-conflict adaptation. *Psychiatry Res.* 185, 149–156.
- Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202.
- Miller, E.K., Erickson, C.A., Desimone, R., 1996. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J. Neurosci.* 16, 5154–5167.
- Miyake, A., Friedman, N.P., 2012. The nature and organization of individual differences in executive functions four general conclusions. *Curr. Dir. Psychol. Sci.* 21, 8–14.
- Morris, N., Jones, D.M., 1990. Memory updating in working memory: The role of the central executive. *British Journal of Psychology* 81, 111–121.
- Moscovitch, M., 2008. The hippocampus as a stupid, domain-specific module: Implications for theories of recent and remote memory, and of imagination. *Can. J. Exp. Psychol.: Rev. Can. DE Psychol. Exp.* 62, 62–79.
- Nadel, L., Moscovitch, M., 1997. Memory consolidation, retrograde amnesia and the hippocampal complex. *Curr. Opin. Neurobiol.* 7, 217–227.
- O'Reilly, R.C., 2006. Biologically based computational models of high-level cognition. *Science* 314, 91–94.
- O'Reilly, R.C., Frank, M.J., 2006. Making working memory work: a computational model of learning in the Frontal cortex and Basal Ganglia. *Neural Comput.* 18, 283–328.
- Oberauer, K., 2001. Removing irrelevant information from working memory: a cognitive aging study with the modified Sternberg task. *J. Exp. Psychol.: Learn. Mem. Cogn.* 27, 948–957.
- Oberauer, K., 2005. Binding and inhibition in working memory: individual and age differences in short-term recognition. *J. Exp. Psychol.: Gen.* 134, 368–387.
- Oberauer, K., Souza, A.S., Druey, M.D., Gade, M., 2013. Analogous mechanisms of selection and updating in declarative and procedural working memory: experiments and a computational model. *Cognitive psychology* 66, 157–211.
- Owen, A.M., McMillan, K.M., Laird, A.R., Bullmore, E., 2005. N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping* 25, 46–59.
- Pelegrina, S., Lechuga, M.T., García-Madruga, J.A., Elosúa, M.R., Macizo, P., Carreiras, M., Fuentes, L.J., Bajo, M.T., 2015. Normative data on the n-back task for children and young adolescents. *Front. Psychol.*, 6.
- Pereg, M., Shahar, N., Meiran, N., 2013. Task switching training effects are mediated by working-memory management. *Intelligence* 41, 467–478.
- Price, J.M., Colflesh, G.J., Cerella, J., Verhaeghen, P., 2014. Making working memory work: the effects of extended practice on focus capacity and the processes of updating, forward access, and random access. *Acta Psychol.* 148, 19–24.
- Rac-Lubashevsky, R., Kessler, Y., 2016. Dissociating controlled and automatic updating in working memory: the reference-back paradigm. *J. Exp. Psychol.: Learn. Mem. Cogn.* 42, 951–969.
- Ragland, J.D., Turetsky, B.I., Gur, R.C., Gunning-Dixon, F., Turner, T., Schroeder, L., Chan, R., Gur, R.E., 2002. Working memory for complex figures: an fMRI comparison of letter and fractal n-back tasks. *Neuropsychology* 16, 370–379.
- Ranganath, A., Jacob, S.N., 2015. Doping the mind dopaminergic modulation of prefrontal cortical cognition. *Neuroscience*, 1–11.
- Rouder, J.N., Morey, R.D., Verhagen, J., Swagman, A.R., Wagenmakers, E.J., 2016. Bayesian Analysis of Factorial Designs. *Psychol Methods*. (In press).
- Schmiedek, F., Li, S.C., Lindenberger, U., 2009. Interference and facilitation in spatial working memory: age-associated differences in lure effects in the n-back paradigm. *Psychol. Aging* 24, 203.
- Schoofs, D., Preuß, D., Wolf, O.T., 2008. Psychosocial stress induces working memory impairments in an n-back paradigm. *Psychoneuroendocrinology* 33, 643–653.
- Segal, A., Kessler, Y., Anholt, G.E., 2015. Updating the emotional content of working memory in social anxiety disorder. *J. Behav. Ther. Exp. Psychiatry* 48, 110–117.
- Sternberg, S., 1966. High-speed scanning in human memory. *Science* 153, 652–654.
- Szmales, A., Verbruggen, F., Vandierendonck, A., Kemps, E., 2011. Control of interference during working memory updating. *J. Exp. Psychol.: Hum. Percept. Perform.* 37, 137–151.
- Tzelgov, J., Henik, A., 1991. Suppression situations in psychological research: definitions, implications, and applications. *Psychol. Bull.* 109, 524–536.
- van Schouwenburg, M., Aarts, E., Cools, R., 2010. Dopaminergic modulation of cognitive control: distinct roles for the prefrontal cortex and the basal ganglia. *Curr. Pharm. Des.* 16, 2026–2032.
- Watter, S., Geffen, G.M., Geffen, L.B., 2001. The n-back as a dual-task: P300 morphology under divided attention. *Psychophysiology* 38, 998–1003.