LUND
UNIVERSITY

Department of Psychology

Volitional and strategic retrieval-orientation adoption in reality monitoring: an ERP study

Nils Nyberg
Bachelor's thesis, spring 2016

Supervisor: Mikael Johansson

Abstract

Retrieval orientations refer to cognitive states that, if adopted, tonically orients episodic retrieval attempts towards the sought-after type of information. Such orientations can enhance episodic retrieval by modulating the processing and effectiveness of retrieval cues. The current study investigated if retrieval orientations are adopted volitionally and spontaneously, as previous studies have not addressed this question. This was investigated in a source-memory paradigm, designed for the present purposes. Participants were exposed to three study-test blocks. At study, participants encoded words denoting everyday objects. Half of these words were followed by a picture of the object, and half by the instruction to imagine the object. At test, two blocks were respectively manipulated so as to make words from one study condition more prevalent, making it beneficial for the participants to adopt an orientation towards this type of encoded item. A third test block acted as a baseline, and did not contain a retrieval-orientation manipulation but instead contained equal amounts of words from the two study conditions. Source accuracy was best for the more prevalent item-type in the two manipulated blocks, and did not differ between the two old item-types in the baseline block. The ERPs from correct rejections were not significantly different between blocks. Moreover, accuracy and ERP analyses conducted on the first versus second half within blocks did not reveal any significant differences. Possible reasons for the obtained results are discussed, such as it not being relevant for participants to adopt retrieval orientations due to their initial high accuracy rate.

Keywords: recognition memory, retrieval orientation, reality monitoring, event-related potentials, multinomial processing tree models

Introduction

The general objective of this study is to advance the current understanding of memory retrieval. More specifically, the present study investigates a cognitive retrieval process, or state, referred to as retrieval orientation. This state, it is thought, can be adopted in order to facilitate retrieval and meet specific retrieval goals. The primary research aim is to explore, using both behavioural and electrophysiological data (in the form of event-related potentials), whether people volitionally and spontaneously adopt such orientations, as this has not been investigated before.

Episodic memory

There exist different types of memories. Some memories involve mentally traveling back in time and reliving past experiences and events. These types of autobiographical memories, conceptualized as episodic memories by Tulving (1972), differ from general fact memories which can involve knowing, for example, that Paris is the capital of France. Tulving termed such fact memories semantic memories, and they function as a person's general-knowledge storage, independent of personal experiences. While both types of memories are long-term declarative memories, episodic memories can reinstate and allow one to remember various phenomenal aspects present during the encoding conditions, such as the emotions felt and the contextual features that surrounded the events (Tulving, 1972).

Episodic memory can be studied in a variety of different tasks. For example, researchers may employ either a task requiring recall of information or a task requiring recognition of information. In a recall task, people are asked to reproduce as much of the previously seen information as they can without the help of retrieval cues. In a recognition task, people are instead asked to make judgements regarding whether or not they have seen an item before, or what group an item was previously presented as being a part of. A recognition task is employed in the current study, as it offers more control over when retrieval cues are shown, and thus when participants retrieval processes begin (Smith & Kosslyn, 2007).

A dual-process model of episodic memory has suggested that recognition memory can be further fractionated into recollection and familiarity (Addante, Ranganath, & Yonelinas, 2012; Mandler, 1980; Yonelinas, 2002). For example, one can recollect detailed information about an acquaintance who is met at random, such as who the person is and where one knows the person from, or the recognition of the person can be based on a sense of familiarity and a sense that one knows the person from somewhere, but without being able to recollect more substantive information regarding the identity of the person.

There is a consensus among memory researchers that episodic retrieval is dependent on an interaction between retrieval cues and neurocognitive memory traces (Smith & Kosslyn, 2007). Retrieval cues can be defined as information or stimuli (external or self-generated) that help to bring memories to conscious awareness and which thus enable access to specific memories in the long-term memory storage. Neurocognitive memory traces refer to the neuronal alterations occurring in especially hippocampus as memories are encoded (Teyler & DiScenna, 1986; Teyler & Rudy, 2007).

Principles of successful episodic retrieval

Successful episodic retrieval depends not just on how similar a retrieval cue is to a neurocognitive memory trace, but also on the cognitive states that have been adopted and on the processes that have been engaged during retrieval. Some of the most well-substantiated retrieval mnemonic principles are described below.

The importance of encoding-retrieval overlap. The "encoding specificity" and "transfer appropriate processing" principles highlight the fact that the same retrieval cue can be more or less effective in activating a memory trace depending on how it recapitulates information or processing that was present at encoding. Thomson and Tulving (1973) introduced the "encoding specificity principle" and they argued that the ease and degree to which a memory trace is aroused depends on the extent to which features of the encoding environment are recapitulated in information and features of the retrieval cue(s). These features can refer to the physical context, social context or the emotional or other psychological state in which the information was encoded and later retrieved in. For example, Weingartner, Adefris, Eich and Murphy (1976) demonstrated that when participants encoded words in an intoxicated state, they had better recall of these words when they were in the same intoxicated state, relative a sober state, during retrieval.

The "transfer appropriate processing" principle, introduced by Morris, Bransford and Franks (1977), states that the type of processing that is involved during retrieval can enhance the chance of retrieval success if it is appropriately matched to the one that was engaged during encoding. The researchers had one group of participants come up with words that rhymed with presented words during encoding, and they had another group of participants focus on the semantic qualities of these words. They then tested both groups of participants on words that rhymed with the original words, and the participants had to recall the original words when seeing these words. They found that participants recall of the original words were improved when they had also focused on words that rhymed with the original words during encoding, as opposed to when they had focused on the semantic qualities of these words.

The role of tonically maintained cognitive states. In 1983, Tulving argued that another prerequisite besides a retrieval cue must be met in order to retrieve episodic information. He argued that individuals need to adopt a specific cognitive state, or mode, termed “retrieval mode”, in order to treat and process either external or internal information as retrieval cues. For example, seeing the word Berlin may or may not bring back memories from a trip to Berlin, depending on the individual’s cognitive state. It is thought that once adopted, this state is maintained for an extended period of time (the state is said to be “tonically” maintained: M. D. Rugg & Wilding, 2000).

More recently, and of central importance to the present study, Rugg and Wilding (2000) proposed that the effectiveness of a retrieval cue depends not just on 1) its similarity to or overlap with a memory trace, or 2) how it recapitulates information or processing at encoding, but also on how the cue is specifically processed during retrieval attempts. They argue that when people want to retrieve specific information, they can orient their memory search towards the targeted information, and this orientation can in turn modulate the specific form of processing that is applied to retrieval cues. For example, if I want to remember where I placed my keys, last seen yesterday, I can adopt a state where I orient and thus bias my memory search towards those memories that were encoded yesterday and which involve the places in my apartment where I normally leave my keys. This state thus makes me process these internal retrieval cues (the keys and my apartment) according to my goal; *finding my keys*, and according to the nature of the memories I want to retrieve; *the place I left my keys in*. These cognitive states have been termed retrieval orientations, and previous research indicates that they states are tonically maintained as well, which is described below.

Previous research on retrieval orientation

EEG/ERP basics. Most studies investigating retrieval orientations have employed the event-related potential (ERP) technique. This technique is based on the electroencephalography (EEG), which are the small continuous voltage fluctuations that originate mainly from postsynaptic potentials (PSPs) in the brain, and which can be picked up by electrodes placed on the scalp (Luck, 2014). Event-related potentials (ERPs) are the voltage fluctuations that are due to specific cognitive, sensory or motor processes. The ERPs usually contains multiple positive and negative peaks, so called ERP components, each of which can potentially be related to a cognitive, sensory or motor process (Luck, 2014). Specific ERPs can be investigated by selecting time-windows of the EEG in which one or more processes of interest is thought to occur, and then averaging these time-windows (within as well as between participants). This cancels out the noise from the other brain, ocular, muscular and electrical-artefactual activity and makes the ERP signal more prominent (Luck, 2014).

Experimental paradigms. Previous research on retrieval orientations have employed a variety of tasks, some of which involve source monitoring, which is the process by which the sources of encoded items have to be retrieved (M. K. Johnson, Hashtroudi, & Lindsay, 1993; Mitchell & Johnson, 2000). Some source-monitoring tasks require that participants, for every item, make a judgement regarding either the source of the item, or if the item is new. This type of task is used in the current study. Another task which can be argued to require source-monitoring is the memory exclusion task, which was created by Jacoby (1991). In the task, participants have to make a target judgement ("old" judgement) towards one type of encoded items at test, while rejecting the other types of old items, along with new items ("new" judgement). This task requires recollection of the source of the targeted item, as both types of items are purported to be familiar to the participants, and basing a judgment solely on whether or not an item seems familiar would therefore not suffice (Jacoby, 1991). Furthermore, some tasks involve participants making old/new judgement for test conditions only containing one type of old item. The old item-type is then changed in different test conditions (see for example: Robb & Rugg, 2002). Even though this task does not require source retrieval, the participants can still adopt an orientation towards the encoded item-type in the different test conditions, in order to judge if the item is old or new.

A special kind of source monitoring is done in reality monitoring, which is the process by which a person attributes a memory to either an internal (self-generated) or an external (perceived) source (M. K. Johnson & Raye, 1981). Only a handful of retrieval orientation studies have used a reality monitoring component, which will be described below. It is thought that two factors affect source and reality monitoring judgements: 1) the memory traces per se, and 2) the later decision process (M. K. Johnson & Raye, 1981). For example, memory traces for stimuli that have been externally perceived include more clear spatial-temporal information and more detailed sensory and semantic attributes, relative the memory traces for internally generated stimuli. The memory traces for internally generated stimuli instead contain more information regarding the amount of attention and the types of cognitive operations that were engaged during encoding, relative memory traces for externally perceived stimuli (M. K. Johnson, 1988; M. K. Johnson & Raye, 1981; Mitchell & Johnson, 2000). It is this information that is retrieved during source retrieval, and which form the basis for source judgements.

Principles for investigating retrieval orientation. Besides bringing to the fore the idea that people can adopt retrieval orientations, Rugg and Wilding (2000) also proposed the principles that must be met in order to find the electrophysiological, hemodynamic or other similar correlates of these orientations. Firstly, they argue that such correlates should be found when contrasting the activity elicited by the same type of retrieval cues when these cues are used to search memory for different types of information. Secondly, they argue that studies should perform contrasts for items in which cue-processing can be separated from retrieval-success processing. This second principle can be met by investigating ERPs that are stimulus-locked to correctly rejected new items (called correct rejections: CR) in test conditions that involve different episodic retrieval demands (such as retrieving pictures versus words). These correctly rejected items can be used for the analysis since a processing of these words do not include confounding effects from retrieval success. Furthermore, if retrieval orientations are in fact tonically maintained, then participants should still be in these cognitive states while processing these words. Evidence that retrieval orientations are indeed tonically maintained is described below. Lastly, Rugg and Wilding (2000) argue that it is important to keep the difficulty of different experimental conditions equal, as potential ERP and behavioural differences might otherwise be due to differences in retrieval effort, and thus due to the cognitive processes that have specifically been recruited to aid retrieval attempts.

Influential retrieval orientation studies. Some of the most influential retrieval orientation studies is presented below. All of these studies followed the principles set out by Rugg and Wilding (2000), and they thus formed ERPs stimulus-locked to CRs.

One of the earliest studies of retrieval orientation was carried out by Rugg and Birch (2000). The researchers examined the ERPs in two test blocks that followed either a “shallow” encoding task (making alphabetical judgements regarding study words) or a “deep” encoding task (incorporating the words into meaningful sentences). Participants made old/new judgements at test, and the researchers found that the ERPs were more positive-going following the shallow encoding task relative the ERPs following the deep encoding task. This difference was found between 300-1400 ms with a maximum over frontal and central sites. This study successfully demonstrated, for the first time, that retrieval orientations influence the processing of retrieval cues even when participants only make old/new judgements.

A year later, Wilding and Nobre (2001) performed an influential retrieval orientation study in which they conducted two experiments. In the first experiment, they had participants retrieve previously encoded phonological or imagery-based information in separate block-conditions (answering either “remember”, “know”, or “old” to items), while in the second experiment they had participants make frequent switches between these tasks within one block. The researchers only found a difference in the first experiment, where the ERP for the condition requiring retrieval of imagery-based information was more negative-going between 500-800 ms, and more positive-going between 1100-1600 ms, relative the ERP for the condition requiring retrieval of phonological information. The effect was found throughout the scalp. This study was the first to suggest that retrieval-orientation adoption does not occur flexibly, which is at odds with the retrieval requirements of everyday life, which often necessitates flexible retrieval of different types of specific information, and thus suggests that retrieval orientations should be adopted flexibly.

Robb and Rugg performed a study in 2002 where they dissociated retrieval-orientation effects from retrieval-effort effects (Robb & Rugg, 2002). They had participants encode, in different study blocks, either pictures or words. Each participant was exposed to two study-test blocks, which were both part of either an easy or a hard condition. For the easy condition, the study-test interval lasted for 30 seconds, during which the participants had to perform a backwards counting task. For the two hard blocks, this interval was five minutes, thus making retrieval harder compared to the easy condition. During test, participants had to judge, depending on the antecedent study block, whether a presented word was the name of a studied picture, or if a word had been shown at study (old/new judgment). They found that ERP effects for the difficulty-manipulation were present until 300 ms post-stimulus, while retrieval-orientation effects were present from 300-1900 ms. The ERPs were more positive-going for the easy condition at frontal sites, but more positive-going for the hard condition at central and parietal sites. The ERP correlates for retrieval orientations were more positive when the study task involved words relative to when the study task involved pictures. The retrieval-orientation effect was found at electrode sites throughout the scalp.

This study was the first to demonstrate that retrieval-effort effects can be separated from retrieval-orientations effects, and moreover that such effort effects can occur earlier than retrieval-orientation effects.

Hornberger, Morcom and Rugg (2004), and Stenberg, Johansson and Rosén (2006) conducted separate studies that manipulated the effect of study-test similarity on retrieval orientations. The first study factorially crossed pictures and words (denoting these pictures) in four study-test cycles (for example, only pictures at study and words at test, or only pictures at study and pictures at test, et cetera). They found that the ERPs elicited by correctly rejected pictures and words were more positive-going when the study material belonged to the same class of studied items for 300-900 ms, with word-cues having a more anterior distribution relative to picture cues. The second study had participants encode both pictures and words (denoting these pictures) in the same study block, and these items could in a later test block appear in the same or the opposite format. One group of participants were instructed to only endorse items that preserved their format from study, while the other group endorsed both formats. The endorsement was done via an old/new judgement. The study confirmed the earlier Hornberger et al. (2004) results, in that the ERPs were overall more positive-going for the group that only endorsed items preserving their format from study (i.e., when there was a study-test similarity) relative the group that endorsed both formats from study. This larger positivity was found throughout the scalp. These results imply that the amount of overlap between study-test items have an impact on the specific orientation-processes engaged. For example, Stenberg et al. (2006) argued that the group endorsing only items preserving their format from study adopted a perceptual retrieval orientation, while the second group adopted a conceptual retrieval orientation.

Lastly, Herron, Evans and Wilding (2016) recently published a study in which they, for the first time, demonstrated that it is possible to find evidence for retrieval orientations even under conditions involving frequent task-switches. At test, participants alternated frequently between three episodic memory tasks, which required either retrieval of one of two different kinds of contextual information for items shown at study (screen location or encoding task) or item recognition (old/new judgement). The recognition task acted as a baseline to compare the ERPs from the other two tasks with. The task-switches were pseudo-randomized in that two retrieval tasks of the same kind were always shown in succession, after which two items requiring another type of retrieval were shown. The researchers argued that previous research using mixed conditions might have failed to detect ERP differences for retrieval orientations since the ERP correlates for these orientations might be smaller in magnitude in mixed conditions, and thus harder to find when more similar items are used. The major difference between this study and the Wilding and Nobre (2001) study is thus that this study purposefully employed retrieval of very different kinds of information. The results revealed that the ERPs for the test condition requiring retrieval of encoding task differed compared to both the ERPs for the baseline and compared to the ERPs for the test condition requiring retrieval of items screen location. The ERPs for the encoding task were more negative-going relative the ERPs for the other tasks between 400-1900 ms, with the largest effect evident at midline sites. The ERPs for the screen location task did not, however, differ from the baseline ERPs and the researchers concluded from this that the retrieval orientation ERP effects were largely driven by the requirement to retrieve encoding operations; a conclusion that could not have been validly drawn without a baseline condition. This study was the first retrieval-orientation study to include a baseline condition, and such an inclusion proved to be valuable as it offered further insights into the differences between the orientations adopted for the different conditions.

Since the presented studies found ERP correlates of retrieval orientations from CRs, this implies that retrieval orientations are indeed tonically maintained cognitive states. The ERP-correlates for retrieval orientations differ in these studies since they depend on the nature of the targeted material, which varied between studies. Furthermore, since these studies hold everything constant in different test conditions except for manipulating the information which participants orient towards, the ERP differences these studies found are evidence for retrieval-orientation adoptions. To summarize, these studies imply that retrieval-orientation effects can be separated from retrieval-effort effects, that task-switching creates problems for finding ERP-correlates of retrieval orientations, and that retrieval-orientation effects can be affected by the similarity between study-phase and test-phase items.

In reality monitoring. There are only two previous studies investigating retrieval orientations in reality monitoring. Leynes, Cairns and Crawford (2005) conducted the first such study. They had participants encode words in two experimental study conditions. In a reality monitoring study condition, participants heard words in a male voice, or they had to generate words themselves. In an external source monitoring study condition, participants heard different words in male and female voices. Participants were then tested in two different source memory task, depending on the preceding study condition. In both of these test conditions, the participants had to choose between one of three responses, one for each source and one for new items. The researchers found more positive ERPs at frontal electrodes between 1000 and 1200 ms for the reality monitoring test-block relative to the external source monitoring test-block. This was the first retrieval orientation study to use reality monitoring, and the first study to show that retrieval-orientation effects in reality monitoring have a characteristic frontal ERP distribution.

The present study is particularly influenced by study from 2011 by Rosburg, Mecklinger and Johansson (Rosburg, Mecklinger, & Johansson, 2011)**.** This study employed a memory exclusion task and the participants had to encode, in the study phase, object names that were followed by either a picture of the object, or by the instruction to imagine a picture of the object. The researchers then tested participants in two different test conditions, and they found that the ERPs were more positive-going at frontal electrodes between 600-1100 ms when imagined items were the target-category relative to when perceived items were the target-category. The researchers also found that reaction time was significantly slower for the imagined items relative the perceived items when both of these items had been targets in different conditions. However, the researchers did not find any differences in reaction time when these items had been non-targets.

This study also gives support to the fact that retrieval of self-generated information seems to more selectively involve frontal areas, relative retrieval of perceptual information. This has, in fact, also been supported by functional magnetic resonance imaging (fMRI) experiments. For example, such experiments have demonstrated that retrieval of self-generated information specifically involves both rostral medial prefrontal cortex (PFC) and left lateral PFC (Lundstrom et al., 2003; Vinogradov, Luks, Schulman, & Simpson, 2008).

The nature of retrieval-orientation adoptions. There is a debate in the literature regarding when retrieval orientations are adopted. One side support the view that participants adopt retrieval orientations when retrieval accuracy is low, and one side support the view that retrieval orientations correlate with high retrieval accuracy. For example, Dzulkifli, Sharpe and Wilding (2004) had participants make two different target judgements for two different types of items in different test conditions. They then calculated the difference between the two types of hit scores and split participants into two groups based on the median of this difference, thus forming a low accuracy group and a high accuracy group. The researchers only found ERP correlates for retrieval orientations for participants in the low accuracy group. However, another study also had participants make two different target judgements for two different types of items in different test conditions, but this study instead divided people in to two groups based on every participants' average hit accuracy, collapsed across the two test conditions (Bridger, Herron, Elward, & Wilding, 2009). These researchers only found ERP correlates for retrieval orientations in participants with a high average accuracy.

Even though the group-split was performed differently for these two studies, they give support to different notions regarding retrieval-orientation adoptions. The first study suggests that retrieval-orientation adoptions are compensatory mechanisms, and thus that orientations are adopted during instances when retrieval is hard. The second study instead suggest that retrieval orientations is a symptom of a more effective usage of retrieval cues, and that these adoptions thus occur in people with high retrieval accuracy (Rosburg et al., 2011).

Research question and motivation for the current study

All previous studies investigating retrieval orientations have used tasks that have in common that the participants have been explicitly told to make a specific judgement for one type, or subset, of the encoded information at test, while rejecting the other type(s) of information. This means that participants have oriented towards the relevant type of encoded information in order to reject or accept test items as belonging to this group. Using such tasks have led to fruitful discoveries regarding the ERP correlates of different retrieval orientations, which has been presented in the previous section. However, the inevitable downside with these tasks is that they have constrained participants' retrieval-orientation adoptions, by specifically instructing and requiring participants to orient towards one type of encoded information, in order to make specific target or old/new judgements. We therefore do not know if people volitionally and spontaneously adopt these orientations, so as to strategically facilitate their retrieval and more aptly meet their retrieval goals. This is the research question the current study aims to investigate and shed light on.

The previous example regarding the missing keys exemplifies the reasonableness in believing that people adopt retrieval orientations wilfully and strategically, and even frequently, in their everyday life. It is also of vital importance that the mechanisms behind such adoptions are investigated and further scrutinized, since it has been suggested that an inability to volitionally and spontaneously adopt retrieval orientations might underlie some of the memory problems associated with both ageing and certain types of brain injury (Hornberger, Rugg, & Henson, 2006).

Since no purposeful paradigm exist to investigate the current research question, an experimental task was designed for the present purposes. It was decided to include a reality monitoring aspect in the form of perceived and imagined objects as study items, as this was successfully implemented by Rosburg et al. (2011), which influenced the current study. During test, old and new words, all referring to objects, are used as retrieval cues. Retrieval-orientation adoptions are actively manipulated by including, in two different test blocks, an unequal amount of words (as retrieval cues) from the two antecedent study conditions. It is hypothesised that it will therefore be beneficial for the participants to orient towards the more prevalent item-type, as this will improve the majority of their retrieval attempts. These manipulated test blocks are henceforth referred to as the active retrieval orientation blocks. The experiment also includes a baseline test block, which does not involve an active manipulation towards making it beneficial for the participants to adopt a retrieval orientation. The non-manipulation in this block is accomplished by having an equal amount of words from the two study conditions.

Having three study-test blocks also allows for the amount of items within each block to be relatively high. This was deemed desirable as it is currently unknown in the literature how many consecutive trials of the same kind are sufficient for participants to understand item contingencies (such as understanding that one type of information is made more prevalent), and thus how many trials are necessary for a retrieval-orientation adoption to occur. However, the literature does describe difficulties in finding evidence for retrieval-orientation adoptions during experiments involving too frequent task-switches. As previously described, Wilding and Nobre (2001) only found ERP correlates of retrieval-orientations for a blocked condition, and not for a mixed condition, and similar results have been found by Herron and Wilding (2006), Johnson and Rugg (2006) and Werkle-Bergner, Mecklinger, Kray, Meyer and Düzel (2005). However, as has been previously described, the recent study from Herron and Wilding (2016) did find ERP differences in conditions with frequent-task switches, by purposefully using highly dissimilar retrieval tasks.

A difference between previous mixed retrieval orientation studies and the current study is that all previous studies use an equal amount of items requiring differential retrieval during test, while the current study includes four times more items from one study condition relative the second study condition. This ratio was subjectively decided upon as it is hypothesised to be sufficiently high enough for participants to understand, unconsciously or consciously, that one type of information is more prevalent, while still entailing a sufficient number of items from the less prevalent condition. The inclusion of less-prevalent items within these blocks ensures that a retrieval-orientation adoption is volitional and spontaneous, since the participants have to choose to orient towards one type of information over the other. Without these items, the argument can be made that the task has inadvertently constrained participants retrieval-orientation adoptions.

As previously mentioned, the middle test block acts as a baseline. The inclusion of a baseline block is an important one, as most previous retrieval orientation studies conducting a pairwise comparison of the ERPs in two different test conditions have not been able to ascribe the ERP differences found to any of these conditions (see for example: Rosburg et al., 2011). This means that it has been impossible to make claims regarding whether the ERP-difference found is due to qualitative differences in the types of processes engaged in the different blocks, or if they are due to quantitative differences in the same processes engaged (Herron et al., 2016). A baseline block thus allows for further investigation into how potential retrieval orientations differ, since the ERPs from both active retrieval orientation blocks are compared to a common baseline block. So far, only Herron et al. (2016), previously described, have employed a baseline during their investigation of retrieval orientations. Thanks to this inclusion, the researcher could make claims regarding what retrieval task was mostly driving the ERP retrieval-orientation effect.

It has been decided for the present study to use a source-monitoring baseline and participants therefore make the same three-way response in this test block as in the other blocks. This type of baseline is chosen as it is believed to allow for a more adequate isolation of potential retrieval orientation ERP-components. This is because a recognition-familiarity baseline (only requiring old/new judgements) differ against the two active retrieval orientation conditions not just in terms of participants' potential retrieval-orientation adoptions, but also in terms of the requirement to retrieve source information in general.

It is hypothesised that participants will come to develop retrieval orientations as a function of time, and thus that it will take x (a currently unknown amount) of trials for the participants to potentially adopt retrieval orientations. Therefore, separate analyses for the first and second half of each test block are done. This allows for an investigation into whether or not there are any overall differences between these halves in terms of retrieval-orientation adoptions.

As will be further described in the methods section, the current study uses a multinomial processing tree model to analyse the accuracy data. Such models offer separate probability estimates for participants' item detectability (i.e., detecting if the item is old or new), source discriminability (detecting the correct source of the item if it is old) as well as various response and guessing biases for items that are either not detected or discriminated (Batchelder & Riefer, 1990).

In conclusion, the current paradigm can capture whether three test blocks, all requiring source memory - two of which are manipulated towards a retrieval-orientation adoption and one which is not - will differ from each other. The paradigm thus allows for initial assessments to focus on whether or not the two active retrieval-orientation blocks differ, respectively, against the baseline block. These orientations can then be further scrutinized and compared in order to investigate if they differ as a function of the item source made most prevalent. These investigations are novel as similar paradigms have not been employed in previous retrieval-orientation studies. The current study thus offers a relevant and important contribution to the literature on retrieval orientations.

Hypotheses

The overall hypotheses are that participants will volitionally and spontaneously adopt two different types of retrieval orientations in the two active retrieval orientation test-blocks. These two adoptions will differ from the retrieval processing in the baseline block, where no adoption will occur. These orientations will also differ as function of the source information made most relevant an as a function of time.

The specific hypotheses, which indicate what results will support these overall hypotheses are as follows: **1)** there will not be any difference in accuracy or reaction time for CR across blocks, since CRs should not be affected by retrieval orientation adoption (even though such an adoption can be measured via ERPs for these trials). **2)** Participants will make more accurate decisions as well as have a faster reaction time for the more prevalent items relative the less prevalent items in the two active retrieval orientation test blocks, indicating that it has been beneficial for them to adopt retrieval orientations. In the baseline block, there will be no significant difference between these items in terms of accuracy or reaction time. **3)** Participants will develop a bias to judge non-discriminated items as belonging to the more prevalent item-type in the two active retrieval orientation blocks, with no such bias developing in the baseline block. This bias will occur as a function of time, and thus be evident when comparing the bias estimates for the first and second halves of the blocks. **4)** The ERPs will be more positive-going at frontal electrode sites for the test block where previously imagined items are more prevalent, relative the test block where previously perceived items are more prevalent [Following Leynes et al. (2005) and Rosburg et al. (2011)]. **5)** The effect from hypothesis four will be more evident in the second half compared to the first half of the two active retrieval orientation blocks. **6)** The ERPs from the two active retrieval orientation blocks will be different from the ERPs for the baseline block (since they will have adopted a retrieval orientation in both the two active retrieval orientation blocks, but not in the baseline block). However, since baseline blocks are not generally included in retrieval orientation studies, it is hard to speculate in the precise manner that they will differ.

Method

Participants

Eighteen right-handed volunteers, 10 females (56 %), that ranged in age from 19-34 years (*M* = 23.44, *SD* = 3.50) participated in the experiment. All of the participants were native Swedish speakers and undergraduate students who were recruited by convenience sampling at mainly Lund University, and to a lesser extent Malmö University College. The recruitment procedure involved entering undergraduate classes and handing out a sheet where people interested in participating could sign up with their name and e-mail address. All participants reported to be of normal health without a history of epilepsy or neurological surgery, and all participants had normal or corrected-to-normal vision. The participants were offered cookies as well as water during two breaks in the experiment, and were compensated with a movie ticket at the end.

Materials

The experiment took place in an EEG lab located in the Department of Psychology at Lund University. For the stimulus material, a total of 450 coloured pictures of everyday objects as well as their names were collected from two stimulus sets: the Bank of Standardized Stimuli (BOSS) (Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010) and the HatField Image Test (Adlington, Laws, & Gale, 2009). The pictures were scrutinized and selected based on two principles set out by Johnson, Kounios and Reeder (1994) and also followed by Rosburg et al. (2011). The first principle was that each picture had to be able to be described by a one-word label (such as “walnut” [In Swedish: “valnöt”]). The second principle was that this word had to be specific and concrete enough that it would evoke a mental image of the object when participants were asked to form such an image after seeing the word (for example, the word "valnöt" does not have any connotations besides its lexical definition as a particular type of nut, and if participants are asked to form a mental image after seeing this word, they will therefore most likely form an image that is a representation of this particular nut).

The items that met these criteria then formed a pool from which the final items were chosen. In this stage, the items were chosen so that they created nine sets, of 50 items each, that were matched for word length and word frequency. The frequencies for each word was determined with the corpus PAROLE which was developed by Språkbanken, a research unit at the Department of Swedish at University of Gothenburg (Ridings & Borin, n. d.). The corpus contains information on approximately 20 million Swedish words collected as part of an EU project that spanned from 1976 to 1997. The words and frequencies in the corpus were collected by Språkbanken from texts in newspapers, magazines, fiction and nonfiction books as well as from webpages.

In the present study the range for word length was 3-10 letters (M = 6.58, SD = 1.98) while the range for word frequency was 2-622 occurrences per million words (M = 53.42, SD = 69,10). The mean word-length in the nine sets was 6.80 (SD = 1.91), 6.52 (SD = 2.02), 6.50 (SD = 1.87), 6.20 (SD = 2.06), 6.64 (SD = 2.01), 6.62 (SD = 1.98), 6.76 (SD = 2.01), 6.42 (SD = 1.97) and 6.76 (SD = 2.11). A one-way ANOVA confirmed that these numbers were not significantly different, $F(8, 441) = .47$, NS. The mean frequency in these sets was 53.52 (SD = 59.10), 53.30 (SD = 93.27), 53.40 (SD = 64.80), 53.80 (SD = 61), 53.12 (SD = 66.20), 53.12 (SD = 74.8), 53.18 (SD = 61.85) 53.46 (SD = 63.64) and 53.88 (SD = 75.88). A one-way ANOVA also confirmed that these numbers were not significantly different, $F(8, 441) < .001$, NS (a p-value below .05 is held to indicate significance, as per standard conventions. The acronym NS stands for not significant).

Design

The experiment was coded in and run using the E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). The instructions as well as the words used for stimuli were written in Arial font, size 34, and shown on a 27-in computer screen with a 2560 ×1440 resolution. The instructions were written in a standard text format (i.e., the letters were lower-case except for first-sentence letters that were written in upper case, et cetera). The words were centred in the middle of the screen and presented in white colour on a black background. The pictures were presented within a 400 × 400-pixel white square centred in the middle of an otherwise black background. The pictures encompassed around two-thirds of this white square (the exact measurements differed somewhat depending on what object the picture in question represented). When the white square was shown without a picture, it followed the same 400 × 400-pixel measurements.

The experiment used a mixed factorial design with two independent variables. One independent variable, the type of test block, was implemented using a within-subject component and contained three levels: perceive block (PB), baseline block (BB) and imagine block (IB). The other independent variable, the ordering of the test blocks, was implemented using a between-subject component and contained two levels: PB first and IB last, and IB first and PB last. This order was counterbalanced across participants, with the restriction that the BB was always in the middle (see Table 1). The motivation for always keeping the baseline block in between the active retrieval orientation blocks is that it keeps the ordering of this test-condition constant for all participants.

Table 1. *The two possible orderings of the three test blocks.*

Order	Test block one	Test block two	Test block three
Order one	Imagine block	Baseline block	Perceive block
Order two	Perceive block	Baseline block	Imagine block

Each participant was exposed to three study-test blocks. The study procedure was always the same and was in large part influenced by the design used by Rosburg et al. (2011). In each study block, participants encoded a total of 80 stimuli, 40 of which were from a perceive item condition (PC) and 40 from an imagine item condition (IC). For both of these conditions, a word referring to an object was first shown. In the PC, a picture of the object was shown after the word, while in the IC a white square was shown after the word, and the participants were asked to imagine and mentally project the object within the white square. The presentation of items was pseudo-randomized, with the constraint that no more than two items from one condition could be shown in succession. Two of the nine stimuli sets were combined to form the study items for each study block; one set was assigned to be used for the PC and one set was assigned to be used for the IC. Since each set contained 50 items, 40 of these items were randomly chosen for each study condition in each study block.

The test blocks were designed specifically for the purpose of this experiment. In each test block, 100 words were shown, but the amount of words taken from the PC and IC (from the antecedent study block) differed in the three test blocks. In the PB, there were 40 words from the PC, 10 words from the IC, and 50 new words. In the BB, there were 25 words each from the PC and the IC and 50 new words. In the IB there were 40 words from the IC, 10 from the PC and 50 new words. When an item category had less than 40 words at test (PC in IB, IC in PB, and both PC and IC in BB), these words were randomly selected from the previously presented 40 words during study. A new (for the participant previously unused), set of items was used for the 50 new items in every test block.

The presentation of words during test was also pseudo-randomized in that no more than three words from one item condition (words from either the PC or IC, or new words) could be shown in succession, and the first and second half of each test block had to contain the same amount of words from the three word conditions.

Repeating the set-assignment procedure for all study-test blocks resulted in all nine sets being used for each participant. Furthermore, all sets were counterbalanced for all item conditions (PC, IC, new words) and block orderings across participants (which required a minimum of 18 participants). During the later data analysis, the data from the PC and the IC were collapsed across block orderings, respectively, so that the data from participants being exposed to a PB first and those being exposed to a PB last were analysed together, and the same with the IB. This analysis-setup was decided upon as it would increase statistical power due there being more data analysed within each condition. This setup also entails that carry-over order effects (such as participants always performing better in the last compared to the first block due to learning effects, et cetera) will affect both the PB and the IB data to the same degree.

Procedure

Participants were randomly selected to participate at one of three time-points throughout the day (at either 9 am, 12 pm or 3 pm), and only one participant conducted the experiment at any one time. The experiment, excluding EEG capping and post-experimental procedures such as debriefing, took between 70-80 minutes. When the participants arrived they were informed about the EEG capping procedure, broadly about the experiment, and that they could at any time and without any need for justification discontinue the experiment. The participants were also told that the results from the experiment were anonymous and that no names or any other information that could be traced back to their person would be used during the analysis and later publishing of the omnibus results. After understanding this information, the participants had to give written consent or they could not continue the experiment (all participants gave written consent). The participants were thereafter fitted with an EEG cap, a procedure that lasted between 40-60 minutes. This procedure consisted mainly of minimizing the electrode impedances by applying a saline solution in the ring-shaped holes of the electrodes placed on the scalp.

After the EEG capping-procedure was completed, the participants were seated in front of a computer screen and told not to make any excessive or unneeded muscular movements. Participants were seated about one meter from this screen. The participants were also told to remain relaxed during the experiment and to try and avoid eye blinks except during the presentation of a question mark in the study phase, or a fixation cross in the test phase (with no restrictions during the instructions). The specific experimental instructions were then shown on the computer screen and read aloud to the participants as they silently read the same instructions on the screen. When the participants had verbally confirmed that they had understood the instructions, the experiment began.

Study-block procedures. Within each study block, each trial began with the presentation of a fixation cross for 500 ms (which the participants had been told to fixate on), after which an object name was presented for 1500 ms. As previously mentioned, a picture of the denoted object was presented in half of the trials (PC), while a white square instead appeared in the other half of the trials (IC). The participants had been instructed before the study procedures began to imagine a picture of the object within the white square when it was shown. At the end of each trial a question mark was shown and the participants had to rate how well the picture (imagined or perceived) illustrated the object in question. These judgements were made by right-hand button presses on the numerical part of a computer keyboard (with 1 = good [bra], 2 = fair [hyfsad] and 3 = bad [dålig]. The words in the brackets designate the original Swedish word-usage). The response alternatives were shown under the question mark. The motivation for including this judgement was that it ensured that participants remained focused on the stimuli throughout the study blocks, and the judgement-data was therefore not analysed. Each trial was subject-terminated by a response, after which the next trial began (see Figure 1, A, for an illustration of the study-block procedures).

Test-block procedures. After a short break that was terminated by the participants (with the constraint that a break had to last at least around 0.5 minutes and could last at most around 3 minutes), a test phase began. The participants were instructed that they were going to see different words and then make a judgement regarding whether a word had, in the previous study block, been followed by a picture that was perceived or imagined, or if the word was new. The participants were also instructed to try and answer as fast and as accurately as possible. Each trial began with a fixation cross shown for 1500 ms, after which an empty black background was shown for 400 ms, before the test words were displayed for 200 ms. A question mark was thereafter shown which indicated that the response interval had started. The response interval lasted for 3800 ms and was not terminated by a response. Furthermore, the response alternatives were not shown under the question mark in this part of the experiment, as the participants had been allowed sufficient time during the test instructions to learn which buttons to press for each response.

The participants were instructed to use the index and second finger of one hand to press certain keys if a test item corresponded to a previously perceived or imagined picture, respectively, and to use either the index or second finger of the other hand to press another key for new words. The mapping between hands and source/new judgements were counterbalanced between participants (the response-buttons were K, L and A or A, S and L, for "perceived" "imagined" and "new" responses, respectively). Lastly, the participants were given accuracy feedback for the preceding five trials, shown after every five trials, which was displayed for 700 ms. The feedback was given in percentage with the word "correct" ("korrekt" in Swedish), and was written in blue text. This feedback was implemented to motivate the participants to do well (see Figure 1, B, for an illustration of the test-block procedures).

The experiment ended with a debriefing which involved informing the participants about the broad motivation and hypotheses for the experiment. The participants were also given the opportunity to ask any remaining questions they had concerning any part of the experiment, or regarding post-experimental data-handling procedures.

Ethical declaration

As has been mentioned, all participants were informed of their right to discontinue their participation at any time without justification, and all participants were told that all subject-related data would be handled and to-be published in an anonymous fashion. The confidentiality and anonymity of the data were ensured by not having any names or other identifiable attributes attached to the data (besides age, gender and handedness, which were not analysed together with the EEG and behavioural data, but were instead used separately to form mean values for these variables and to ensure that all participants were right-handed). When the data was not handled by a researcher within the memory group at the Department of Psychology, Lund University, it was saved on a USB stick that was kept safe in a locked apartment, and as backup-files on a password-protected cloud-saving service.

The experiment did not involve any deceptive elements or any other components that could pose a potential threat or harm to the participants. The participants were informed that the procedures for attaching the EEG cap could cause some discomfort, but that it should not hurt. The participants were encouraged to tell the experimenters if they during any part of this procedure, or during the experiment as a whole, felt any discomfort, at which point the current procedure would have been stopped.

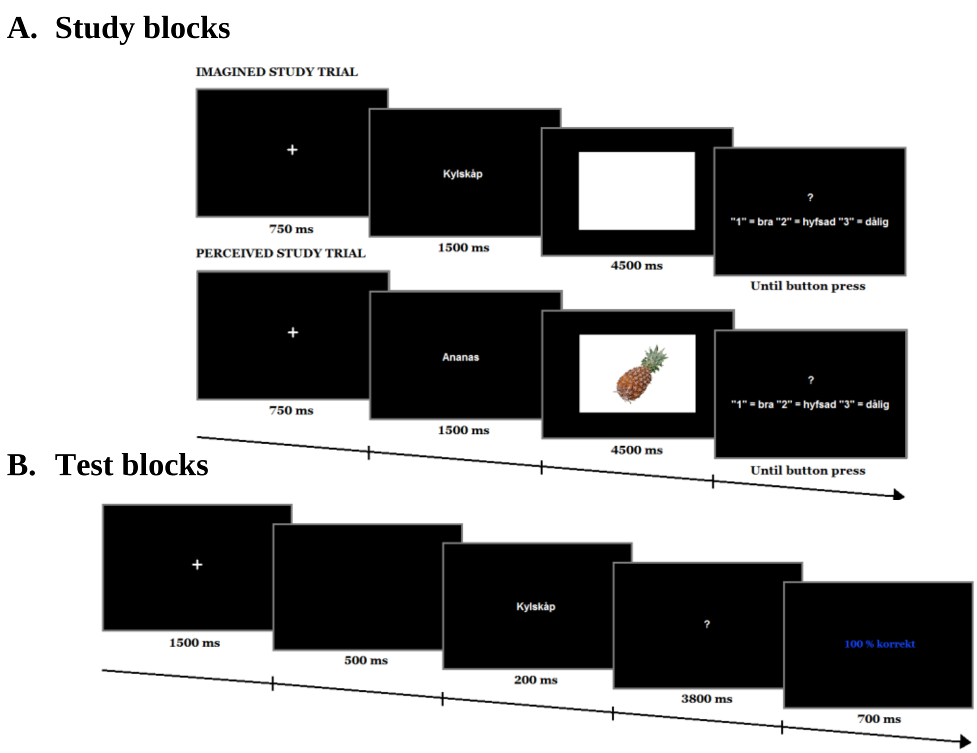

Figure 1. Study and test block trial procedures.
A. An Illustrates of study-block procedures for every trial. Separate illustrations are found for the imagined study conditions (IC, above) and for the perceive study conditions (PC: below). B. An illustration of the test-block procedures for every trial. The feedback (the last display) was only shown after every 5 trials, and gave accuracy feedback for the last five trials. The arrow in both A and B indicate that the ordering of the different parts of the trials should be read from left to right.

Data processing and analysis

EEG/ERP acquisition and pre-processing. The study used an elastic EEG cap (Easycap, Herrsching-Breitbrunn, Germany) that contained 31 Ag/AgCl-sintered ring electrodes. Electrooculogram (EOG) electrodes were attached to the temple of each lateral canthus as well as below the left eye, and reference electrodes were attached to the left and right mastoid, with the left mastoid functioning as the active recording reference. The ground electrode was attached to the cap, in position AFz. Electrode locations were based on the 10-20 system. The EEG data was sampled online at 2000 Hz and amplified to 200 Hz, using a Neuroscan NuAmps EEG/ERP amplification system. This data was later down-sampled offline to 500 Hz. For each participant, the impedances for the main electrodes were kept below 5 kΩ, the impedances for the EOG electrodes below 15 kΩ, and the impedances for the mastoid electrodes below 3 kΩ.

The EEG data was pre-processed using the FieldTrip MATLAB toolbox for EEG/MEG-analysis (Oostenveld, Fries, Maris, & Schoffelen, 2010: Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands). Firstly, an initial high-pass filter of 0.01 Hz was applied to the data, and the data was re-referenced to the average of the two mastoids. The horizontal channels and the vertical EOG (VEOG) channel together with the Fp1 channel were also re-referenced offline to form, respectively, bipolar horizontal EOG (HEOG) and VEOG channels. EEG epochs were then extracted, stimulus-locked to the onset of CRs in the three test blocks. The epochs extended from -200 ms to 2000 ms after the onset of these words. The pre-stimulus interval was used for baseline correction. Each epoch was thereafter visually inspected and only epochs demonstrating severe abnormal trends were rejected at this stage (the rejection was thus liberal). An independent component analysis (ICA) was then applied to the data, and components containing VEOG, HEOG and in some instances muscle and electrical-noise artefacts were corrected for. After this step, the epochs were again inspected and a more conservative epoch rejection was done (i.e., epochs with EEG activity exceeding ±100 μV and epochs exhibiting severe abnormal trends were excluded). A final low-pass filter of 30 Hz was then applied to the data. After these steps, separate ERP averages were formed for the CRs in the three blocks, and ERP averages were also formed for the CRs in the first and second half within each block. This procedure was first done separately for each participant, after which grand average ERPs were formed for all the 18 participants.

For the grand averages, the amount of CR trials used from each participant in the PB ranged from 41-49 (M = 46.6, SD = 2.3), with the range of trials for the first and second half ranging from 19-25 (M = 23.5, SD = 1.5) and 20-25 (M = 23.1, SD = 1.3), respectively. The amount of CR trials used for each participant in the IB ranged from 39-50 (M = 46.8, SD = 2.9), with the range of trials for the first and second half ranging from 20-25 (M = 23.5, SD = 1.6) and 18-25 (M = 23.3, SD = 2), respectively. The amount of CR trials used for each participant in the BB ranged from 35-50 (M = 45.6, SD = 3.5), with the range of trials for the first and second half ranging from 20-25 (M = 23.3, SD = 1.5) and 14-25 (M = 22.3, SD = 2.7), respectively.

ERP data analysis. The mean amplitude was chosen as an ERP comparison measure in the current study. The time windows chosen were based on a visual-inspection of the ERP data, and will be described in the results section. The initial repeated measures ANOVA included the factors CR BLOCK (PB, BB, IB), LOCATION (anterior, central, posterior) and HEMISPHERE (left, midline, right). A follow-up repeated measures ANOVA was thereafter conducted to investigate if the ERP correlates for potential retrieval orientations differed as a function of time. This ANOVA therefore included the above-mentioned factors, as well a BLOCK HALF (first, second) factor. Only effects involving CR BLOCK or BLOCK HALF will be described. A total of 14 electrodes were included in the analysis (Fp1, Fp2, F7, F3, F4, F8, FC5, FC6, CP5, CP1, CP2, CP6, P3, P4).

Behavioural data analysis.

Accuracy. The behavioural accuracy data was analysed with a multinomial processing tree model. These models use mathematical parametrization to come up with probability estimates for the underlying cognitive processes hypothesised to be engaged during source retrieval (Batchelder & Riefer, 1990; Riefer & Batchelder, 1991). Such cognitive processes involve item-status detection (if the item is old or new), source discrimination, as well as various response and guessing biases. These models can be illustrated as tree diagrams. The “roots” (which are found on the leftmost side of the trees in Figure 2) refer to either an old item or a new item, the branches refer to conditional parameter probabilities and the “leaves”, or end notes, refer to the judgement that the participant makes in regard to the item. In the current experiment, the items were words, and each word could be judged as belonging to either a previous perceived condition (PC), a previous imagined condition (IC), or the word could be judged to be new. As seen in Figure 2, there were, for every type of word, several branches that combined into the same source judgement. There were three trees within each half of each block, with 15 ending-branches, that led to a total of nine separate judgements (“PC”, “IC” or “New” judgements for the three types of words).

The parameters D_1 and D_2 were the probabilities of correctly detecting PC and IC words as old, respectively, while 1-D_x (where x = 1, or 2) were the probability that these words are not detected as old. If a participant correctly identified an item as old, (s)he could either remember or have forgotten the source of the item. The parameters d_1 and d_2 were the probabilities of correctly judging the source of item PC and IC, respectively, while 1-d_x (where x = 1, or 2) were the probability that the source was not detected. The other parameters found in the trees referred to various response and guessing biases. The bias to respond “old” to undetected items was demonstrated by parameter *b*, while the bias to respond new to these items was reflected by 1-*b*. In cases when an item had successfully been detected, but the source not discriminated, parameter *a* referred to the probability of guessing that the item came from source PC, and 1-*a* similarly reflected the probability that participants would guess that an item came from source IC (Batchelder & Riefer, 1990).

Multinomial processing models are analysed with goodness-of-fit tests, which are computed with the asymptotically χ^2 distributed log-likelihood ratio statistic G^2 ((Batchelder & Riefer, 1990). Hypothesis tests are performed both to test the fit of different models against the empirically obtained data, and to conduct hypothesis tests between models. Certain constraints often have to be put on a model, both for it to fit the data, and to manage degrees of freedom so that further tests can be performed with the chosen model.

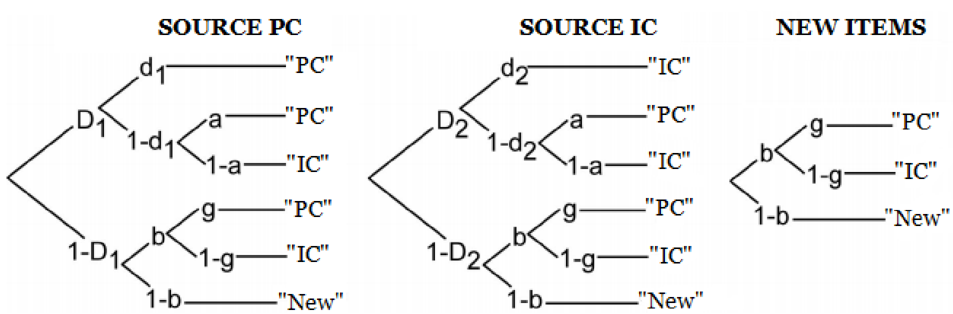

Figure 2. Multinomial processing tree models for source PC, IC and new items.
The illustrations depict three multinomial processing tree models which are used to measure item detectability, source discrimination, as well as response/guessing biases. The model has a total of seven parameters for each source, and four parameters for new items. D_1 = detectability of PC words, D_2 = detectability of IC words, d_1 = source discriminability for PC words, d_2 = source discriminability for IC words, *a* = bias to respond “PC” to non-discriminated words, *b* = bias to respond “old” to a non-detected words; *g* = guessing that a non-detected word belongs to source “PC”. A total of 18 trees was built, with separate trees made for the first and second half for each test block, within each test block. PC = Perceive condition, IC = imagine condition.

Batchelder and Riefer, (1990) describe seven different models, each with different constraints. These models are obtained by equating either the detection parameters (D1 = D2), the discrimination parameters (d1 = d2), and/or the guessing parameters (a = g). The b parameter is assumed to not differ between any conditions in any of these models. The selected model should describe and thus fit the data, but the model should not be so constrained that the results cannot be generalized. According to Heck, Moshagen and Erdfelder (2014), models should reflect the principle referred to as Occam's razor, which in this case means that if a more constrained model can fit for the data equally well as a model with less constraints, then the more constrained (and thus simpler) model should be chosen.

To test a models fit, the goodness-of-fit test is conducted on the data, and the test in this case measures the distance between the category-frequencies implied by the model and the category frequencies based on the sample data (Batchelder & Riefer, 1990; Riefer & Batchelder, 1991). The null hypothesis is that the model fits this data, meaning that significant results indicate a bad fit. Once a model has been chosen, hypothesis tests between models can be computed, by comparing the fit of a model to another nested model that has certain parameter constraints (meaning that more parameters have either been set to equal each other, or some parameters have been set to equal a certain number). Hypothesis tests are computed with a ΔG^2 statistic and the null hypothesis states that there is no difference in fit between the models (the new model fits the data equally well as the original model). The research hypothesis thus state that these models do not fit the data equally well, and that there is a significant difference between the parameters in the original model and the way they have been constrained in the nested model.

To obtain the probabilities that participants make certain source judgements, the branches in the trees have to be multiplied. For example, the probability of answering "PC" for a word belonging to the IC can be obtained by equation 1, while the probability of answering "IC" for a word belonging to the PC can be obtained by equation 2.

1. $Pr("PC"|IC) = D2(l - d2)a + (1 - D2)bg$
2. $Pr("IC"|PC) = D1(l - d1)(1 - a) + (1 - D1)b(1 - g)$

The java-based freeware MultiTree (Moshagen, 2010: University of Mannheim), was used to both construct the multinomial tree models, and to conduct parameter estimation and hypothesis tests on models. A total of 18 trees was built, one for each word type, for each half within each block. The result of the model-testing procedure is described in the results section. The PC and IC items were first analysed and compared separately for each block, (the two halves within each block were collapsed to form overall data) before first versus second half analyses were performed.

Since multinomial processing models do not offer an easy way to compare CRs across blocks, an initial repeated ANOVA was conducted on the CR accuracy data, which included the factors BLOCK CR (PB, BB, IB) and BLOCK HALF (first, second).

Reaction time. An initial ANOVA was performed on the CR reaction time data, to investigate whether or not reaction time for these words differed across blocks and/or halves. This ANOVA included the factors BLOCK CR (PB, BB, IB) and BLOCK HALF (first, second). Reaction time analyses were thereafter performed separately for every block-type, with a repeated-measures ANOVA using the factors ITEM TYPE (PC, IC) and BLOCK HALF (first, second).

When Mauchly's test of sphericity had been violated for any behavioural or ERP data, the degrees of freedom and *p*-values were corrected for using Greenhouse-Geisser estimates (Greenhouse & Geisser, 1959). For post-hoc paired-samples t-tests, Cohen's *d* was calculated using the formula $t_{c,}$ presented in Dunlap, Cortina, Vaslow and Burke (1996). A *p* value below .05 was always held to indicate significance. However, where pairwise comparisons were made (such as for pairwise comparisons of nested models for the accuracy data), the *p* values were compared to an adjusted alpha value, corrected for using the Bonferroni-Holm procedure (Holm, 1979). This means that the alpha level was, in practice, sometimes lower than .05. All presented *p* values were below this alpha value unless indicated with the acronym NS, which stands for not significant.

Results

Behavioural results

Accuracy. See the previous Methods section for an explanation of the data analysis. For a table of the accuracy results, see Table 2.

Correct rejections. The CR result did not demonstrate either a significant BLOCK CR main effect [$F(2, 34) = .331$, NS], a significant HALVES main effect [$F(1, 17) < .001$, NS], or a significant BLOCK × HALVES interaction effect, $F(2, 34) = .047$, NS. This creates the ideal settings for investigating the ERP differences, as these results do not suggest that there were any differences between blocks in terms of their difficulties or in terms of the effort participants recruited.

Choosing a multinomial processing tree model. An initial data fit was tested on a model where all of the detectability parameters (D_1 and D_2 in each block) were set to equal D_1 in the first block and the bias and guessing estimates (a and g in each block) were set to equal g in the first block. This model fit the data well $G^2(9) = 13.03$, $p = .161$, which indicated that there were no significant differences between the parameters that were set to equal each other, either within or across blocks. There is thus no difference for any detection parameters within or between blocks, and no significant difference for response and guessing biases within and between blocks. A model where these parameters were constrained in the above-mentioned fashion was therefore retained and used to further test whether the source-memory parameters d_1 and d_2 generally differed from each other. All parameters were set to equal d_1 in the first block and the results indicated that this model did not fit the data well $G^2(14) = 59.42$, $p < .001$. This implied that these parameters generally differed, either within or across blocks, or both. The model used for the further analyses therefore allowed these items to vary freely, and is equal to model 5b in Batchelder and Riefer (1990).

Source accuracy as a function of block. For the PB, the results revealed that source memory was better for the PC ($M = 0.92$, $SD = 0.06$) relative the IC ($M = 0.86$, $SD = 0.14$), $\Delta G^2(1) = 23.61$, $p < .001$. For the IB, the results revealed that that source memory was better for the IC ($M = 0.94$, $SD = 0.05$) relative the PC ($M = 0.84$, $SD = 0.20$), $\Delta G^2(1) = 6.37$, $p = .01$. For the BB, the results indicated that source memory was not significantly different for the two types of items, $\Delta G^2(1) = 1.03$, NS.

First and second half analyses. For the first vs second half analyses, an initial data fit was tested on a model where all the D_x parameters (where x = 1, or 2) were set to equal D_1 in the first half of the first block, where all the a parameters were set to equal the g parameter in the first half of the first block, and where the remaining g parameters were also set to equal the first g parameter. The results indicated that this model fit the data well, implying that there was no difference between the parameters set to equal each other across blocks or block halves $G^2(21) = 17.41$, NS. There is thus no difference for any detection parameters between the block halves, and no significant difference for response and guessing biases between the block halves. This model was thus retained and a model where the second d_1 and d_2 parameters were set to equal the first d_1 and d_2 parameter within each block, respectively for every block, was compared with this model. The results indicated that there was no significant difference between these models, implying that there was no difference in source memory for the PC and the IC words between the first and second half of each block, $\Delta G^2(6) = 8.52$, NS.

Table 2. *Accuracy data for PC, IC and new items in the PB, the BB and the IB.*

	PC hits			CRs			IC hits		
Block	**Total**	1st half	2nd half	**Total**	1st half	2nd half	**Total**	1st half	2nd half
PB	**0.92**	0.92	0.93	**0.97**	0.97	0.97	**0.86**	0.87	0.84
	(0.06)	(0.08)	(0.05)	**(0.04)**	(0.05)	(0.03)	**(0.14)**	(0.18)	(0.20)
BB	**0.93**	0.91	0.94	**0.97**	0.97	0.97	**0.94**	0.94	0.93
	(0.08)	(0.08)	(0.07)	**(0.04)**	(0.05)	(0.03)	**(0.06)**	(0.08)	(0.10)
IB	**0.84**	0.89	0.80	**0.98**	0.98	0.98	**0.94**	0.93	0.95
	(0.20)	(0.14)	(0.24)	**(0.03)**	(0.05)	(0.03)	**(0.05)**	(0.07)	(0.06)

Note. Mean accuracy data for every word type in every block type. Separate measures are displayed for first and second block halves. The numbers stand for percentages. Standard deviations are shown in parentheses. PB = Perceive block, BB = baseline block, IB = baseline block. PC = perceive word condition IC = imagine word condition, CR = correctly rejected new words.

Table 3. *Reaction time data for PC, IC and new items in the PB, the BB and the IB.*

	PC hits			CRs			IC hits		
Block	**Total**	1st half	2nd half	**Total**	1st half	2nd half	**Total**	1st half	2nd half
PB	**1157**	1197	1117	**936**	969	903.4	**1246**	1408	1295
	(327)	(357)	(298)	**(329)**	(362)	(306)	**(419)**	(450)	(478)
BB	**1166**	1168	1163	**946**	942	950.4	**1243**	1310	1176
	(299)	(271)	(332)	**(260)**	(256)	(272.)	**(340)**	(345)	(331)
IB	**1329**	1332.7	1325.7	**982**	1008.7	956.2	**1194**	1270.5	1118.1
	(396)	(391)	(412)	**(284)**	(269)	(303)	**(316)**	(335)	(285)

Note. Mean reaction time data for every word type in every block type. Separate measures are displayed for first and second block halves. The numbers stand for milliseconds. Standard deviations are shown in parentheses. PB = Perceive block, BB = baseline block, IB = baseline block. PC = perceive word condition IC = imagine word condition, CR = correctly rejected new items.

Reaction time. For an explanation of factors used in the ANOVAs, see the Methods section. For a table of the reaction time results, see Table 3.

Correct Rejections. Firstly, the ANOVA for CRs did not demonstrate a significant BLOCK CR main effect, $F(1.46, 24.89) = .924$, NS [with the assumption of sphericity violated, $W = .634$, $\chi^2(2) = 7.29$, $p = .026$]. Furthermore, the result did not show a significant main effect of BLOCK HALF [$F(1, 17) = 5.17$, NS], or a significant BLOCK CR × BLOCK HALF interaction effect [$F(2, 34) = .953$, NS]. This mimics the accuracy CR data and thus also provides ideal conditions for the later ERP analysis.

Perceive block. The results revealed a significant main effect of ITEM TYPE [$F(1, 17) = 12.22$, $p = .003$, $\eta_p^2 = .418$], indicating that the mean RT for the IC ($M = 1246$, $SD = 419$) was significantly slower than the RT for the PC ($M = 1157$, $SD = 327$).

Baseline block. The results revealed a significant main effect of ITEM TYPE [$F(1, 17) = 5.23$, $p = .035$, $\eta p^2 = .235$], indicating that the mean RT for the IC ($M = 1243$, $SD = 340$) was significantly slower than the RT for the PC ($M = 1166$, $SD = 299$). The analysis also revealed a significant ITEM TYPE × ITEM HALF interaction, $F(1, 17) = 7.57$, $p = .014$, $\eta_p^2 = .308$. Follow-up post-hoc t-tests indicated that this was due to the response time for IC being significantly faster in the second half ($M = 1176$, $SD = 331$) compared to the first half ($M = 1310$, $SD = 345$), $t_{(17)} = 3.08$, $p = .007$, $d = .45$.

Imagine block. The results revealed a significant main effect of ITEM TYPE [$F(1 , 17) = 12.80$, $p = .002$, $\eta_p^2 = .430$], indicating that the RT for the PC ($M = 1329$, $SD = 396$) was significantly slower than the RT for the IC ($M = 1194$, $SD = 316$).

ERP data.

Four time-windows were chosen for the ERP analyses, based on a visual inspection of where the ERPs for the different blocks diverged. These time-windows were 500-700 ms, 1100-1300 ms, 1300-1500 ms as well as 1900-2000 ms. For an illustration of the ERP-plots from six electrodes, se Figure 3.

500-700 ms time window. The ANOVA did not reveal any significant main or interaction effect involving CR BLOCK between 500-700 ms, $F(2/4, 34/68) < 2$, NS for all analyses involving CR BLOCK. Further analyses did not reveal any significant main or interaction effects involving

BLOCK HALF, F(1/2/4, 17/34/68) < 2, NS for all effects involving CR BLOCK and/or BLOCK HALF.

1100-1300 ms time window. The ANOVA did not reveal any significant main or interaction effect involving CR BLOCK between 1100-1300 ms, F(2/4, 34/68) < 2, NS for all analyses involving CR BLOCK. Further analyses did not reveal any significant main or interaction effects involving BLOCK HALF, F(1/2/4, 17/34/68) < 2, NS for all effects involving CR BLOCK and/or BLOCK HALF.

1300-1500 ms time window. The ANOVA did not reveal any significant main or interaction effect involving CR BLOCK between 1300-1500 ms, F(2/4, 34/68) < 2, NS for all analyses involving CR BLOCK. Further analyses did not reveal any significant main or interaction effects involving BLOCK HALF, F(1/2/4, 17/34/68) < 2, NS for all effects involving CR BLOCK and/or BLOCK HALF.

1900-2000 ms time window. The ANOVA did not reveal any significant main or interaction effect involving CR BLOCK between 1900-2000 ms, F(2/4, 34/68) < 2, NS for all analyses involving CR BLOCK. Further analyses did not reveal any significant main or interaction effects involving BLOCK HALF, F(1/2/4, 17/34/68) < 2, NS for all effects involving CR BLOCK and/or BLOCK HALF.

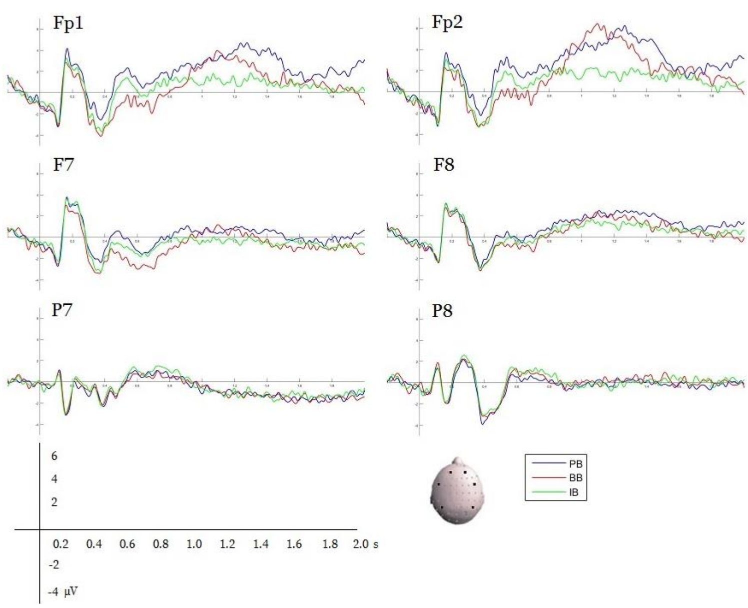

Figure 3. ERP plots for six electrodes.
The grand average ERP plots for six electrodes are shown (Fp1, Fp2, F7, F8, P7, P8). The ERP waves are formed from correct rejections. The x-axis displays seconds, while the y-axis displays microvolts. The head on the left symbolize where the electrodes are located on the scalp. PB = Perceive block, BB = baseline block, IB = imagined block.

Discussion

Interpreting the results.

The purpose of the present study has been to investigate if people volitionally and spontaneously adopt retrieval orientations, in order to strategically facilitate their retrieval. Since no previous study had investigated this question, a paradigm was designed explicitly for this purpose. It was decided to make one type of item from study four times more prevalent than the second item-type, in two different active retrieval orientation test blocks. Accuracy (including detection, discrimination, response-bias and guessing estimates) and reaction-time analyses has been carried out on the hit and CR data in each block, and separate analyses has been conducted for the first and second halves of these blocks. The ERP analyses has been performed by comparing the CR ERPs from every test block against each other, and separate analyses has also included the first and second halves as factors.

The accuracy data mostly support the first and second hypotheses. Hypothesis **1)** states that there will not be any difference in accuracy or reaction time for CRs across blocks, and no such difference is found. Hypothesis **2)** states that participants will make more accurate source decisions as well as have a faster reaction time for the more prevalent items relative the less prevalent items in the two active retrieval orientation test blocks, with no significant difference in these measures for old items in the baseline block. The results revealed this to be mostly true, with the exception that reaction time for the IC is significantly slower than the reaction time for the PC in the BB. The results also reveal a significant interaction effect between block type and item half, which is caused by the reaction time for the IC being significantly faster in the second half of the BB compared to the first half. Despite there being a difference in reaction time for the two old items in the BB, there is no difference in accuracy for these items within this block, or among its halves.

Even though this result did not follow the hypothesis, Rosburg et al. (2011) also found that reaction time for words belonging to a previous IC was slower than words belonging to a previous PC, but only when comparing conditions where both of these words had respectively been targets, and not when comparing conditions where these words had respectively been non-targets. Participants in Rosburg et al (2011) did not make source judgements, but instead had to respond “old” only to one type of old words (the targets), and “new” to the second type of old words (the non-targets) along with new words. In the current study, participants made source judgements for the old words, which means that participants had to retrieve the source for every type of old word in order to make a correct judgement.

Combining the result from Rosburg et al. (2011) and the current study suggests that when items from a PC and an IC are compared in situations where these words are evenly presented, and in situations that require some form of target or source judgement, people can faster retrieve items from a PC relative items from an IC. This also explains why the current study only found a difference in reaction time for these conditions in the baseline block. The other blocks contain more words from one of these conditions, due to the active retrieval-orientation manipulations, and they are thus not presented evenly. A possible reason for why participants can faster retrieve information that belongs to a PC is that the memory traces for perceived stimuli contain more sensory and semantic attributes relative the memory traces for imagined stimuli, and these attributes might be recalled and evaluated fast during retrieval attempts.

It is unclear why participants have a faster reaction time in the second half of the BB for IC words compared to the first half, while such an improvement was not found for these words in the other blocks, or for PC or new words. However, a possible explanation has to do with the BB always being in the middle (between the two active retrieval orientation blocks). For example, participants might evaluate what source every old word belongs to for a long time in the first block, and especially for IC items, as these items might be harder to retrieve (as suggested above). This evaluation can become quicker in the BB, as participants become more confident and faster to retrieve information, and this increase in evaluation efficiency might then peak in the second half of this block. It is also reasonable that this potential learning effect and/or increase in confidence does not affect the words from the PC to the same extent, since these words might not have been as hard to initially retrieve as words from the IC.

Hypothesis **3)** states that participants will develop a bias to answer that non-discriminated items belong to the more prevalent item-type in the two active retrieval orientation blocks, and that this bias will develop over time and thus be more evident in the second half compared to the first half of these blocks. However, no difference was found for any response or guessing bias within or between blocks, or between the halves of any block.

The last three hypotheses concern the ERP results, and the hypotheses are that: **4)** the ERPs will to be more positive-going at frontal electrode-sites for the IB relative the PB, **5)** that this effect will be more prevalent in the second half compared to the first half of these blocks, and that **6)** both the IB and PB ERPs will differ from the BB ERPs. No support for any of these hypotheses were found when analysing the data, since there are no significant main or interaction effects involving CR block or block half.

The electrophysiological data therefore does not support the hypothesis that participants adopted different retrieval orientations in the two active retrieval orientation test blocks, since there are no differences between the ERPs for these blocks. This is a surprising result given that the ERP plots, prima facie, seem to indicate that there are large differences between blocks (see Figure 3). Furthermore, this data prima facie also seem to support the fact that this effect is largest at frontal electrodes, which is in line with previous studies (Leynes et al., 2005; Rosburg et al., 2011).

It is at this point interesting to note that even though the electrophysiological data does not give support to the conclusion that retrieval-orientation adoptions occurred, the accuracy and reaction-time results follow some of the same patterns that would be expected for such adoptions. Namely, participants have better accuracy and faster reaction times for the more prevalent items in the two active retrieval orientation test blocks, which seems to indicate that participants oriented towards these items and benefited from it. However, since participants did not adopt a bias to respond to the more prevalent item-type, this further complicates the interpretation of these results.

There are primarily two explanations for why some of the behavioural results give support to the fact that retrieval-orientation adoptions occurred, while the ERP results do not give support for such adoptions. One explanation for these results is that participants did adopt different retrieval orientations, but that this was not “picked up” by the electrophysiological data. Based on the literature, there can be at least three possible reasons for this.

Firstly, the items used for the different study and test conditions might be too similar for the adopted retrieval orientations to display significantly different ERP waves, which is in line with the suggestion by Herron et al. (2016) that states that similar item-types might create similar ERP correlates for the retrieval orientations. However, it is unlikely that this explains the ERP results, since both Leynes et al. (2005) and Rosburg et al. (2011) found ERP differences for a reality-monitoring retrieval-orientation paradigm with items that were as similar between each other as those used in the current study. Furthermore, even though both types of studied items are pictures, either perceived or imagined, the memory traces for these different types of pictures are highly dissimilar. As previously mentioned, perceived pictures have more sensory and semantic attributes in their memory traces, while imagined pictures have more information regarding the amount of attention and the cognitive operations that were engaged during encoding in their memory traces. Since this information is suggested to be retrieved during source judgements, the retrieval-processing occurring for perceived and imagined pictures can be concluded to be highly dissimilar. This means that potential retrieval orientations should also be dissimilar in the ERP data, as such orientations are tonically maintained extensions of retrieval processes.

Secondly, since participants had to switch between retrieval tasks within blocks, the ERP correlates for the potential retrieval orientations may not be as distinct, due to electrophysiological "noise" from these switches, such as occurs if an adoption is disrupted with every switch. In support of this explanation, Herron et al. (2016) suggests that the ERP correlates for retrieval orientations might be smaller in magnitude in mixed conditions. This means that it cannot be ruled out that retrieval-orientation adoptions occurred in previous mixed-block studies, even though these studies did not find ERP results to support such conclusions. If the retrieval-orientation ERP-correlates are small in magnitude in the current study, but nonetheless exist in the ERP, it can be argued that there should still be a difference when comparing the ERPs for the two active retrieval-orientation test blocks with the ERP for the baseline block. This is because the baseline block is not manipulated towards any retrieval-orientation adoption, meaning that retrieval-orientation ERP-correlates should be non-existent in any magnitude in this block. Since no difference is found for the ERPs between the three blocks, this either means that the ERP correlates for retrieval orientations are so small that they simply cannot be found significant between any blocks, or that they did not occur at all in the ERP. This second possibility will be further described below.

It is also possible that task-switches leads people to alternate between adopting and not adopting retrieval orientations. These potential transient retrieval-orientation adoptions might have occurred at different times for different participants, which means that periods of non-adoptions for one participant can cancel out periods of adoptions for another participant when ERP averages are created across participants. If this is true, this has decreased the power in the current study to find retrieval-orientation ERP-correlates.

Thirdly, since no ERP differences could be found for the test blocks, even though the ERPs for these blocks look highly distinctive at especially frontal areas, this indicates that there was a lot of electrophysiological variance between participants. This variance might indicate that some participants adopted retrieval orientations while others did not. The ERP retrieval-orientation effects might thus, again, be cancelled out due to the fact that people who did not adopt such orientations are included in the averages.

If participants did adopt retrieval orientations, then why is this not evident in the bias scores? One reason that participants did not develop any particular biases is that their accuracy was overall very high, meaning that there were very few times participants were uncertain of the source of a word. This also means that any potential biases will be harder to find statistically, since there are few wrong judgements to base these statistical analyses on. The potential problem of participants' high accuracy will be further discussed below.

Another explanation for why the ERP result indicated that no retrieval-orientation adoptions occurred, is that no such adoptions did occur. There are primarily two possible reasons for this. Firstly, as has previously been discussed, most previous studies employing mixed conditions have not been able to find ERP correlates for retrieval orientations. In order to combat this problem, the current study included four times more words from one study condition relative the second study condition in two active retrieval orientation blocks. It was therefore hypothesised that task-switches would not affect the current study. However, since no previous retrieval-orientation study has used a paradigm similar to the current one, it cannot be ruled out that the task-switches disrupted participants from adopting any retrieval orientations at all. This might be due to the task-switches occurring too frequently, which did not allow participants to "build up" a retrieval orientation.

The second reason for why participants might not have adopted retrieval orientations is that their accuracy is very high even from the beginning. In the PB and the IB, the judgement accuracy for the most prevalent word-type as well as new words is above 90 % while accuracy for the less prevalent word-type is above 80 %, for both halves. For the BB, all words are correctly judged above 90 % of the time, for both halves. Since there is no significant difference in accuracy between halves in any block, the argument can be made that a selective orientation towards one type of old information does not benefit participants, since they already make highly accurate judgements. As previously mentioned, it is thought that retrieval orientations are adopted in order to more aptly meet specific retrieval goals, but if retrieval goals are met before such adoptions have occurred, a strong argument can be made that there is no reason to alter the current retrieval strategy. This might mean that support for a volitional and spontaneous retrieval-orientation adoption can only be found in conditions where retrieval is harder than in the current task. This would support the notion that retrieval-orientation adoptions occur as compensatory mechanisms, during times when the sough-after material cannot be easily retrieved.

However, if participants did not adopt orientations, then what is a possible explanation for participants having higher accuracy for the more prevalent items in the PB and IB? A possible explanation is that participants, either unconsciously or consciously, expect or become accustomed to retrieve information from the more prevalent item-type, without tonically orienting towards this kind of information. Participants higher accuracy for the more prevalent items might thus be due to procedural memory-mechanisms, meaning that since participants are more used to retrieving information from the more prevalent item-type, they become more sensitised towards retrieving this information.

Conclusion. The conclusion to be drawn from the above interpretation is that the results are inconclusive in regard to whether or not participants volitionally and spontaneously adopted retrieval orientations. Some of the behavioural results seem to indicate that participants did adopt different retrieval orientations, while some of the behavioural results and all of the ERP results does not support this conclusion. The present study can therefore neither fully discard or give support to the initial research question. However, and as described in the introduction, since people often have to flexibly adapt to changing episodic retrieval-demands in everyday life, it is reasonable to believe that people do volitionally and spontaneously adopt strategic retrieval orientations, and the initial research question should therefore be further researched in future studies. Suggestions for how these studies should be carried out will be described in later sections.

Methodological strengths and weaknesses.

The present study has various methodological strengths and weaknesses, some of which, it will be discussed, can have affected the current results.

Strengths. Since the present study is the first to investigate the current research question, a paradigm was specifically designed with the purpose of finding volitional and spontaneous retrieval-orientation adoptions. Since it is thought that retrieval orientations are adopted in a strategic manner at least during some episodic retrieval attempts in everyday life, the current paradigm mimics more of the real-world characteristics of such adoptions relative previous retrieval orientation studies. The current study therefore has high ecological validity relative these previous studies, since these studies have in common that they have constrained participants' retrieval-orientation adoptions, which does not occur in most real-life situations.

The present study can also be argued to have a high statistical-conclusion validity, especially regarding the behavioural-accuracy results. These results were analysed with a multinomial processing tree model and, as has been described, the advantages of such a model is that it offers separate measures for the various cognitive processes and bias-judgements proposed to be engaged during source retrieval. Many standard statistical accuracy-measures do not allow for separate estimations of these cognitive processes and biases, meaning that they often become entangled within the same statistical measure. The present study instead separately investigated if item detectability, source discriminability as well as response and guessing biases differed for the words from the different types of sources and new words. Multinomial processing tree models can therefore be argued to be more appropriate for source-memory accuracy-analyses, as they allow for more comprehensive assessments of source data.

The current study not only used behavioural data, but complemented this data with electrophysiological measures in the form of ERPs. This has added additional validity to the data-interpretation, as an interpretation focusing on a combination of behavioural and ERP data offers more insights than an interpretation focusing only on one of these measures. Since EEG measures brain activity at the millisecond level, this has made it possible to investigate if support for retrieval orientations can be found in the ERP data, and exactly when such potential orientations affect retrieval-cue processing. The behavioural data, in turn, has allowed for judgements to be made regarding whether or not such adoptions is beneficial for participants' accuracy and reaction time. An investigating based solely on one of these measures by themselves would have afforded a conclusion based on less information, which means that the conclusion would be more prone to be wrong.

Another strength of the current study is the inclusion of the baseline block. Even though no significant ERP results were found, the baseline block allowed for further insight into the possible reasons for the obtained results (see above). A strength of the current study is also that this study is the first to measure if people volitionally and spontaneously adopt retrieval orientations, and it has thus been the first study to develop a paradigm to investigate this question. Future studies can hence learn from the current approach and the obtained results when creating new studies.

Weaknesses. This section will describe potential weaknesses in the current study, and suggestions for improvements will be presented in later sections. A weakness of the current study is the many, albeit argued for, subjectively decided-upon elements. These elements include: the amount of study-test blocks, the amount of items within these blocks, and the ratio between the more prevalent and less prevalent words within the PB and the IB. It is unknown what effect the choices made for the current study have on the obtained results, but it is possible that these parameters are not the most ideal ones for finding evidence for volitional and spontaneous retrieval-orientation adoptions. To investigate what the ideal parameters are, further research is needed.

Another weakness of the current study is that participants who might have adopted retrieval orientations were analysed together with participants who might not have adopted retrieval orientations. It is also highly unlikely that participants adopt a retrieval orientation exactly midway through a test block, and the crude first versus second half analyses therefore run the risk of clumping together trials where no adoptions occur with trials where adoptions occur, thus decreasing the power to find evidence for retrieval orientations.

There is also a problem of construct validity in the current study due to the fact that it is unknown if the current study-manipulations are hard enough, or give participants enough motivation, to adopt retrieval orientations, and therefore if volitional and spontaneous retrieval-orientation adoptions is what the current experiment is measuring.

Another methodological weakness of the current study is one inherent in ERP analyses generally. When pre-processing ERP data, trials are subjectively excluded by visual inspections, and the decision regarding what components to keep and remove from an ICA is also based on the researchers' subjective judgements. It is therefore possible that the choices made by the researchers in the current study can have an effect on how the ERP data turned out.

Lastly, this study suffers from the same potential problem that most psychology and cognitive neuroscience experiments suffers from, namely that the study only used undergraduate students as participants. There is therefore a potential problem of external validity, as this sample may not represent the population as a whole, and the results obtained may therefore not align with the results that would be obtain from a more representative sample. Likewise, the number of participants (18) has been chosen as it is the minimum number required for all the experimental conditions to be counterbalanced. However, it is possible that a larger sample would have led to different result, and thus that the non-significant current results are due to the small number of participants.

Suggestions for future studies.

Based on the results in the current study, many suggestions can be made for future studies that will investigate if people volitionally and spontaneously adopt retrieval orientations, and the conditions in which this might occur. Firstly, future studies should investigate if more support for such orientations can be found in tasks where retrieval is harder than in the current study. It has been discussed that retrieval-orientation adoption is as compensatory mechanisms, occurring under conditions when retrieval is hard. Since it is possible that the current retrieval tasks are not hard enough for participants to adopt retrieval orientations, there are many possible ways to make similar tasks harder in future studies. For example, future studies can include more items at both the study and test phases, thus making it harder for participants to accurately remember the source of items presented at test. Future studies can also decrease the time participants have to encode study items. However, it is important that studies do not create too hard encoding tasks, since they then run the risk of participants not encoding the items, which in turn can hinder future retrieval-orientation adoptions. Furthermore, following the experimental setup of Robb and Rugg (2002), future studies can increase the time interval between study and test blocks, or the time interval between study and test blocks can include some kind of task, such as backwards counting, as this is also hypothesised to make retrieval harder.

Another discussed explanation for the current results is that the task-switches might have disrupted participants' retrieval orientations, and future studies should therefore investigate if a larger ratio between less prevalent and more prevalent items has an effect on the results. However, if studies use very few or no items at all for a less prevalent condition they run the risk that the retrieval-orientations they find have been constrained by the task. This is because the current research question demands that orientations are adopted when participants choose, unconsciously or consciously, to orient towards one type of information over other types of information.

Future studies should also use more dissimilar item-types for the different study conditions than the current study used. It has been argued that it is unlikely that the chosen item-types affected the current results in a negative way, but it could not be concluded that this was the case and a definitive answer can thus only be obtained by more research.

Another suggestion for future studies is that they too should include a baseline condition, as both the current study and Herron et al. (2016) have successfully demonstrated that baselines offer further insights when interpreting the data, regardless of whether or not the results are in line with the hypotheses.

There are also many different and possible better ways to analyse the data. For example, future studies should in addition to averaging the ERP-data across all participants, also perform separate investigations for high and low performers. Such analyses can be carried out, for example, by conducting a median or average split on the accuracy data, thus forming groups of high and low performers. Such analyses have the potential to be very fruitful, as the current data suggests that some participants might have adopted retrieval orientations while others did not, and separating participants based on their accuracy can potentially allow these differences in adoption to be found in the ERP data. Previously presented studies have also shown that group splits allows for further insights to be gained into the nature of retrieval-orientation adoptions (i.e., if orientations are compensatory mechanisms or symptoms of more effective usage of retrieval cues: Bridger et al., 2009; Dzulkifli et al., 2004).

Lastly, conducting first versus second half analyses of the test blocks is not optimal, since the first half might contain trials where participants adopted retrieval orientations or the second half might contain trials where participants did not adopt retrieval orientations. The block-data should therefore, if enough trials are included, be split into more than just two groups that are based on the block halves, as this will allow for a further investigation to be made regarding when in time retrieval orientations are adopted.

Implications of the current and future studies.

The current results imply that further research is needed in order to investigate both if people spontaneously and volitionally adopt retrieval orientations, and under what conditions such orientations occur. Since the current results were inconclusive, the study has not been able to generate new knowledge regarding whether or not people volitionally and spontaneously adopt retrieval orientations, but the current study has been able to generate knowledge for future studies, as the current paradigm and results can lead to more improved future paradigms. The current study has also been able to further demonstrate the importance of a baseline in retrieval-orientation studies.

Many important implications can potentially come from future studies investigating the current research question. As mentioned in the introduction, it has been suggested that aging as well as certain types of brain injury directly affect the ability to volitionally and spontaneously adopt retrieval orientations. Therefore, if the mechanisms behind such adoptions are further scrutinized, it is possible that the results from these future studies can aid the development of more specialized treatments.

Furthermore, it is possible that future studies investigating the current research question can help shed light on the possibly interaction or correlation between retrieval-orientation abilities and other psychological abilities. For example, it can be hypothesised that there is a correlation between being able to volitionally and spontaneously adopt a retrieval orientation and being able to keep one's attention without being distracted.

Acknowledgements

I would like to thank Casper Kerrén for all his help and support throughout the project. I could not have coded my experiment in e-prime without his help. I would also like to thank Anna Karlsson for her help with capping my participants, Inês Bramao for demonstrating how to use the FieldTrip MATLAB toolbox and the Memory Group for all their helpful suggestions. Lastly, I would also like to thank my supervisor, Mikael Johansson, for his guidance and advice.

References

Addante, R. J., Ranganath, C., & Yonelinas, A. P. (2012). Examining ERP correlates of recognition memory: Evidence of accurate source recognition without recollection. *Neuroimage, 62*(1), 439-450.

Adlington, R. L., Laws, K. R., & Gale, T. M. (2009). The Hatfield Image Test (HIT): A new picture test and norms for experimental and clinical use. *Journal of clinical and experimental neuropsychology*, *31*(6), 731-753.

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological review*, *97*(4), 548.

Bridger, E. K., Herron, J. E., Elward, R. L., & Wilding, E. L. (2009). Neural correlates of individual differences in strategic retrieval processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(5), 1175.

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PloS one*, *5*(5), e10773.

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.

Dzulkifli, M. A., Sharpe, H. L., & Wilding, E. L. (2004). Separating item-related electrophysiological indices of retrieval effort and retrieval orientation. *Brain and Cognition, 55*(3), 433-443.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*(2), 95-112.

Heck, D. W., Moshagen, M., & Erdfelder, E. (2014). Model selection by minimum description length: Lower-bound sample sizes for the Fisher information approximation. *Journal of Mathematical Psychology*, *60*, 29-34.

Herron, J. E., Evans, L. H., & Wilding, E. L. (2016). Electrophysiological evidence for flexible goal-directed cue processing during episodic retrieval. *Neuroimage, 132,* 24-31.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65-70.

Hornberger, M., Morcom, A. M., & Rugg, M. D. (2004). Neural correlates of retrieval orientation: effects of study–test similarity. *Journal of cognitive neuroscience, 16*(7), 1196-1210.

Hornberger, M., Rugg, M. D., & Henson, R. N. (2006). fMRI correlates of retrieval orientation. *Neuropsychologia, 44*(8), 1425-1436.

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

Johnson, J. D., & Rugg, M. D. (2006). Electrophysiological correlates of retrieval processing: Effects of consistent versus inconsistent retrieval demands. *Journal of cognitive neuroscience, 18*(9), 1531-1544.

Johnson, M. K. (1988). Reality monitoring: An experimental phenomenological approach. *Journal of experimental Psychology: general, 117*(4), 390.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological bulletin, 114*(1), 3.

Johnson, M. K., Kounios, J., & Reeder, J. A. (1994). Time-course studies of reality monitoring and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(6), 1409.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological review, 88*(1), 67.

Leynes, P. A., Cairns, A., & Crawford, J. T. (2005). Event-related potentials indicate that reality monitoring differs from external source monitoring. *The American journal of psychology*, 497-524.

Luck, S. J. (2014). *An introduction to the event-related potential technique*: MIT press.

Lundstrom, B. N., Petersson, K. M., Andersson, J., Johansson, M., Fransson, P., & Ingvar, M. (2003). Isolating the retrieval of imagined pictures during episodic memory: activation of the left precuneus and left prefrontal cortex. *Neuroimage, 20*(4), 1934-1943.

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

Mitchell, K. J., & Johnson, M. K. (2000). Source monitoring: Attributing mental experiences. *The Oxford handbook of memory*, 179-195.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of verbal learning and verbal behavior, 16*(5), 519-533.

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods, 42*(1), 42-54.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2010). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience, 2011*.

Ridings, D., & Borin, L. (n. d.). PAROLE-korpusen vid Språkbanken. Retrieved 16 January, 2016, from University of Gothenburg, http://spraakbanken.gu.se/resource/parole

Riefer, D. M., & Batchelder, W. H. (1991). Statistical inference for multinomial processing tree models *Mathematical psychology* (pp. 313-335): Springer.

Robb, W. G., & Rugg, M. D. (2002). Electrophysiological dissociation of retrieval orientation and retrieval effort. *Psychonomic Bulletin & Review, 9*(3), 583-589.

Rosburg, T., Mecklinger, A., & Johansson, M. (2011). Electrophysiological correlates of retrieval orientation in reality monitoring. *Neuroimage, 54*(4), 3076-3084. doi: 10.1016/j.neuroimage.2010.10.068

Rugg, M. D., Allan, K., & Birch, C. S. (2000). Electrophysiological evidence for the modulation of retrieval orientation by depth of study processing. *Cognitive Neuroscience, Journal of, 12*(4), 664-678.

Rugg, M. D., & Wilding, E. L. (2000). Retrieval processing and episodic memory. *Trends Cogn Sci, 4*(3), 108-115.

Smith, E. E., & Kosslyn, S. M. (2007). *Cognitive Psychology: Mind and Brain*. Upper Saddle River, N.J: Pearson/Prentice Hall.

Stenberg, G., Johansson, M., & Rosén, I. (2006). Conceptual and perceptual memory: Retrieval orientations reflected in event-related potentials. *Acta psychologica, 122*(2), 174-205.

Teyler, T. J., & DiScenna, P. (1986). The hippocampal memory indexing theory. *Behavioral neuroscience, 100*(2), 147.

Teyler, T. J., & Rudy, J. W. (2007). The hippocampal indexing theory and episodic memory: updating the index. *Hippocampus, 17*(12), 1158-1169.

Tulving, E. (1972). Episodic and semantic memory. *Organization of Memory. London: Academic, 381*(4).

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological review, 80*(5), 352.

Weingartner, H., Adefris, W., Eich, J. E., & Murphy, D. L. (1976). Encoding-imagery specificity in alcohol state-dependent learning. *J Exp Psychol Hum Learn, 2*(1), 83-87.

Werkle-Bergner, M., Mecklinger, A., Kray, J., Meyer, P., & Düzel, E. (2005). The control of memory retrieval: Insights from event-related potentials. *Cognitive Brain Research, 24*(3), 599-614.

Wilding, E. L., & Nobre, A. (2001). Task-switching and memory retrieval processing: Electrophysiological evidence. *Neuroreport, 12*(16), 3613-3617.

Vinogradov, S., Luks, T. L., Schulman, B. J., & Simpson, G. V. (2008). Deficit in a neural correlate of reality monitoring in schizophrenia patients. *Cerebral Cortex, 18*(11), 2532-2539.

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.