Olfactory working memory: Exploring the differences in *n*-back memory for high and low verbalisable odorants

Andrew Moss, Christopher Miles, Jane Elsley & Andrew J. Johnson*

Department of Psychology, Faculty of Science & Technology, Bournemouth University, Poole, UK

*Correspondence should be sent to: Andrew Johnson, Department of Psychology, Bournemouth University, Poole House, Talbot Campus, Fern Barrow, Poole, UK.

Abstract

We describe four experiments each examining *n*-back performance for high and low verbalisable odorants. Participants were presented with a sequence of odorants and were required to state if the current odorant was the same or different to the odorant presented two items earlier. Experiment 1 reported superior performance for high, relative to low, verbalisable odorants and was evident despite above chance memory performance for the low verbalisable odorants. Experiment 2 showed that such superiority persisted with a concurrent articulation condition, suggesting that the memory benefit was not a consequence of verbal recording and rehearsal. Experiment 3 employed metacognitive judgments and showed that correct 2-back responses for high verbalisable odorants received more recollection responses compared to low verbalisable odorants. Experiment 4 compared *n*-back performance across different stimulus types and showed that, for high verbalisable odorants, performance correlated with both letters and abstract shapes, but such correlations were absent for low verbalisable odorants. Taken together, these findings show differences in *n*-back performance between high and low verbalisable odorants, and show that high verbalisable odorants exhibit performance similarities with both verbal and visual stimuli. We further argue that *n*-back performance for low verbalisable odorants operates differently to that of high verbalisable odorants.

199 words

Keywords: Olfaction; working memory; *n*-back; verbalisation; metacognition

Introduction

The current series of studies examine olfactory working memory using the *n*-back task. In the *n*-back task participants are presented with a continuous sequence of items and, following the presentation of each new item in the sequence, are required to decide whether the current item is the same as the item presented *n* trials previous. The task is conceptualised as a measure of working memory[1] since performance of the task requires both maintenance and updating of the preceding *n* items in memory (e.g. Kane, Conway, Miura, & Colflesh, 2007).

There are a number of proposed processes by which the *n*-back task can be performed. Indeed, *n*-back strategy can vary both between participants and as a function of task demands (Botvinick, Braver, Barch, Carter, & Cohen, 2001). One proposal, referred to as a high control strategy, involves rolling rehearsal whereby the last *n* items in the sequence are rehearsed and dynamically updated within a dedicated rehearsal window (e.g. Juvina & Taatgen, 2007). A second proposal suggests that for some stimulus types it may not be possible to 'rehearse' the last *n* items in memory, and without maintenance within a dynamically updated rehearsal window, participants rely upon item 'familiarity' to perform the task (referred to as a low control strategy, Juvina & Taatgen, 2007; Nijboer, Borst, van Rijn, & Taatgen, 2016). Specifically, it is proposed that when the test item evokes a strong feeling of experimental familiarity, the item is a probable candidate for the target (i.e. the item presented *n*-trials earlier). However, if that test item exhibits low levels of experimental familiarity, it is unlikely that it was presented *n* items earlier.

---

[1] Debate exists with respect to the extent to which the *n*-back task is a genuine measure of working memory. This is beyond the scope of the present study (for more detail see, for example, Jaeggi, Buschkuehl, Perrig & Meier, 2010; Kane et al., 2007; Redick & Lindsay, 2013).

Although the extent to which *n*-back working memory processes/strategies are equivalent across different stimulus types is unclear, correlational studies have shown some relationship across different modality *n*-back tasks. For example, Schmiedek, Lövdén, and Lindenberger (2014) compared 3-back verbal (visual presentation of digits) and 3-back visuo-spatial (dots presented within a 4x4 grid) performance and reported a strong positive correlation ($r = .66$). Jaeggi, Buschkuehl, Perrig, and Meier (2010) compared auditory-verbal *n*-back and visuo-spatial *n*-back and found that the magnitude of the performance accuracy correlations were influenced by memory load and stimulus type. For 1-back, 2-back, and 3-back tasks, the magnitude of cross-stimulus correlation varied from 0.11 to 0.62. For the 3-back comparisons, the correlations were 0.18 (Experiment 1) and 0.25 (Experiment 3) and it is not clear why these correlations should differ so starkly to those reported by Schmiedek et al. (2014). This disparity may be explained by methodological differences across these studies (e.g. differences in stimulus set size and presentation modality of verbal stimuli), however, this may speak to the debate existing with respect to the general reliability of the *n*-back task (Jaeggi et al., 2010; Kane et al., 2007; Oberauer, 2005; Redick & Lindsey, 2013; Schmiedek, Hildebrandt, Lövdén, Lindenberger, & Wilhelm, 2009; Schmiedek et al., 2014; Unsworth & Spillers, 2010). However, reliability is argued to be highest in the *n*-back task when $n>1$ (Friedman et al., 2008; Jaeggi et al., 2010; Kane et al., 2007; Shelton, Elliott, Hill, Calamia, & Gouvier, 2009), and therefore judgments where $n>1$ will be used in the present set of experiments.

That there is shared variance between different modality *n*-back tasks suggests common mechanisms underpinning performance of the task. This suggestion is supported by imaging studies indicating common brain regions (within the prefrontal cortex) engaged in performance of the *n*-back task. For example, using fMRI, Nystrom, Braver, Sabb, Delgado, Noll, & Cohen (2000) found similar patterns of activation (i.e. bilaterally in premotor, supplementary motor, anterior cingulate, superior and inferior parietal, and superior, middle, and inferior frontal areas) for performance of the *n*-back task with verbal, spatial, and non-verbal visual stimuli, and this pattern of activation was affected analogously by changes in memory load. Similarly, common regions (including the dorsal and ventral prefrontal cortex, inferior parietal cortex, Broca's area, insula, anterior cingulate, and premotor areas) were activated during *n*-back for both letters fractals (Ragland et al., 2002). Furthermore, a meta-analysis of imaging research demonstrated that modality-independent brain regions (both frontal and parietal) are employed during performance of the *n*-back task (Owen, McMillan, Laird, & Bullmore, 2005). This analysis included a range of stimulus types (e.g. verbal and non-verbal) that required identity or spatial judgments. Owen et al. reported robust activation in both the dorsolateral and ventrolateral prefrontal cortex, irrespective of stimulus type.

Notwithstanding the above findings, due to the limited research examining olfactory *n*-back performance, it is unclear to what extent *n*-back memory for odours shares such commonalities. To date, only two published studies focus on *n*-back memory for olfactory stimuli (Dade, Zatorre, Evans, & Jones-Gotman, 2001; Jönsson, Møller, & Olsson, 2011). In the olfactory *n*-back task participants are presented with a continuous sequence of odorants and are required to judge whether the current odorant matches the odorant presented *n* items earlier. Both previous studies identified that participants are capable of performing olfactory 2-back tasks (Dade et al., 2001; Jönsson et al., 2011) and the effect of odorant verbalisability on *n*-back performance was examined by Jönsson et al. They presented participants with a continuous 36-item sequence comprising odorants previously identified in pilot work as either high or low verbalisable (Experiment 1). They showed that whilst correct responses for low verbalisable odorants were above chance, performance for high verbalisable odorants was superior (proportion correct = .61 and .77, respectively; an effect replicated in Experiment 2 using a high/low verbalisability blocked design). The present research extends the work of Jönsson et al. by further exploring the performance difference between high and low verbalisable odorants in the *n*-back task and, additionally, explores the extent to which performance on the olfactory *n*-back mirrors that for other stimulus types.

The findings of Jönsson et al. (2011) illustrate that the potential to verbally label the odorants is important in facilitating olfactory *n*-back performance. Whilst there exist only two studies examining olfactory *n*-back performance, there are a range of olfactory memory studies that may also utilise working memory resources. A number of these studies have examined the importance of odorant verbalisability in both short- and long-term memory. For example, a range of studies (see below) have attempted to minimise the participant's ability to verbally re-code the odorants in order to examine the extent to which explicit olfactory memory is possible in the absence of verbal recoding. Whilst removing all semantic information and verbal labelling associated with an odorant is likely impossible (White, Møller, Köster, Eichenbaum, & Linster, 2015), attempts have been made to minimise verbal re-coding via both diligent odorant selection and the employment of concurrent tasks. Consider, for example, Olsson, Lundgren, Soares, and Johnansson (2009) who categorised odorants as either identifiable or unidentifiable dependent upon the participants' ability to generate the veridical verbal label for each odorant. Whilst yes/no recognition for identifiable odorants was superior, recognition for unidentifiable odorants was above chance. Similarly, Møller, Wulff, and Köster (2004) employed odorants that were unfamiliar and, therefore arguably, hard-to-name (see Moss, Miles, Elsley, & Johnson, 2016), and reported above chance performance on both implicit and explicit memory tasks. These studies provide evidence supporting the existence of an olfactory memory system that operates independently of a verbal re-coding process (although it remains possible that some rudimentary verbal labelling may have been used to support memory).

Verbal re-coding in olfactory memory may also be restricted via the employment of a secondary verbal task. Andrade and Donaldson (2007) examined yes/no recognition for odorants where, in one condition, participants performed an additional verbal task (memory for a 7-digit sequence) during the retention interval. Recognition performance was above chance and, of theoretical consequence, was significantly higher than the condition in which the secondary task employed olfactory stimuli. This finding was taken as evidence for the greater use of a perceptual (i.e. the odorant) rather than verbal (i.e. a label for the odorant) representation underpinning recognition. This is consistent with research showing that substitution errors in olfactory memory tend to reflect the perceptual similarity of the target odorant rather than the semantic similarity of the odorant labels (White, Hornung, Kurtz, Treisman, & Sheehe, 1998). Other studies have employed concurrent articulation (CA) as an irrelevant secondary verbal task (e.g. Miles & Hodder, 2005; Reed, 2000). CA acts to disable the articulatory rehearsal process (e.g. Baddeley, Lewis, & Vallar, 1984; Baddeley, Thomson, & Buchanan, 1975; Cocchini, Logie, Della Sala, MacPherson, & Baddeley, 2002) and, thereby, disrupts the ability to verbally recode the odorants and/or rehearse verbal labels. Whilst Miles and Hodder (2005) reported a damaging effect of CA on a 2-alternative forced choice (2AFC) recognition task for 5-odorant sequences, recognition memory with CA remained well above chance (similar trends were apparent in Reed, 2000; see also Annett & Leslie, 1996). The aforementioned reviewed studies all minimise verbal re-coding by either odorant selection or through the use of a secondary task. These studies all show above chance performance, and, taken together, therefore suggest that olfactory memory is not uniquely reliant upon verbal recoding. Indeed, olfactory identification is typically very poor (Jönsson & Olsson, 2003; Jönsson, Tchekhova, Lönner, & Olsson, 2005), with ‘tip-of-the-tongue’ states (strong feeling-of-knowing without a name) accompanied with very little semantic information (Jönsson & Olsson, 2003; Jönsson et al., 2005). That individuals are poor at odorant identification adds weight to the proposition that olfactory memory can operate without verbal labels.

Whilst above chance memory performance remains for olfactory stimuli in conditions where verbal re-coding and rehearsal is restricted, a number of studies have shown that the availability of a verbal label can facilitate memory. For example, verbal labelling of odorants improves both short-term (Frank, Rybalsky, Brearton, & Mannea, 2011; Jehl, Royet, & Holley, 1997; in contrast cf. Zucco, 2003) and long-term odorant recognition (Jehl et al., 1997; Kärnekull, Jönsson, Willander, Sikström, & Larsson, 2015; Lyman and McDaniel, 1986; although for opposing findings see Engen & Ross, 1973), and such verbal labels are more effective when semantically meaningful (as opposed to an arbitrary label given to an odorant during an experiment, Frank et al., 2011; Jehl et al., 1997). That verbal re-coding improves memory for odorants is, perhaps, unsurprising because the verbal label provides additional information about the odorant (Andrade & Donaldson, 2007; Annett & Leslie, 1996; Paivio, 1990; Stevenson & Mahmut, 2013a). This additional verbal information can produce a dual (label and percept) memory trace (Paivio, 1990), or, as suggested by others (Herz & Engen, 1996; White et al., 2015) can result in the odorant now being represented in memory as a verbal label rather than an olfactory percept.

One interpretation of the quantitative differences in memory performance between high and low verbalisable odorants (e.g. Jönsson et al., 2011; Olsson et al., 2009) is that the mental representations for these two stimulus types are qualitatively different. Such an interpretation is supported by imaging data showing that, for a delayed-match-to-sample working memory task, nameable odorants are associated with sustained activity in the inferior frontal gyrus, whereas in contrast, hard-to-name odorants are associated with activation in the piriform cortex (Zelano, Montag, Khan, & Sobel, 2009). Recent work from our laboratory suggests that this neurological dissociation between nameable and hard-to-name odorants may also be observed behaviourally (Moss, Miles, Elsley, & Johnson, 2018). Using odorants previously categorised as either high or low verbalisable (Moss et al., 2016), Moss et al. (2018) demonstrated strong evidence for proactive interference (PI) with high verbalisable odorants, that was absent with low verbalisable odorants. Given that PI effects are evident for both verbal and visual stimuli (e.g. Craig, Berman, Jonides, & Lustig, 2013; Monsell, 1978; Moss et al., 2018), this may be taken to suggest that memory representations for low verbalisable odorants are qualitatively different to those for other stimulus types. Indeed, qualitative differences in the memorial representations for high and low verbalisable odorants may provide an explanation for the contradictory serial position functions reported for olfactory stimuli. Specifically, in some studies odorants have produced serial position functions qualitatively consistent with those found for other types of stimuli (e.g. Johnson & Miles, 2007; Miles & Hodder, 2005; Miles & Jenkins, 2000; Moss et al., 2018; White & Treisman, 1997). However, in other studies, memory for odorants have produced serial position curves that are qualitatively different to other stimulus types (e.g. Johnson, Cauchi, & Miles, 2013; Johnson & Miles, 2009; Moss et al., 2018; Reed, 2000). It is entirely plausible that such performance differences arise because of variations in the use of high and low verbalisable odorants across studies. Based upon our previous PI findings (Moss et al., 2018), we suggest that it is the high verbalisable odorants that produce memory functions similar to those produced by other stimulus types and the low verbalisable odorants do not.

That memory functions for high verbalisable odorants are similar to those for other stimulus types (verbal and visual) is supported by the olfactory *n*-back task conducted by Dade et al. (2001). Here, participants were presented with either a continuous sequence of 12 faces or 12 scented puffs of air and were required to state whether each stimulus was the same as, or different to, the stimulus experienced two trials earlier (i.e. 2-back task). Using positron emission tomography (PET) scans, Dade et al. found that *n*-back for both faces and scents engaged the dorsolateral and ventrolateral prefrontal cortex. The authors argued that this finding supported the idea that working memory processes are independent of stimulus modality. However, as argued by Jönsson et al. (2011), it is worth noting that Dade et al. used familiar odorants (peach, geraniol, eucalyptus oil, costus oil, patchouli oil, and cinnamon-bark oil) that are purportedly easy to name. Consequently, any claims that the neural activation for the olfactory *n*-back matches that for the visual *n*-back is confined to high verbalisable odorants. Indeed, as noted above, the behavioural data of Moss et al. (2018) suggests memory for low verbalisable odorants is in fact different to that of high verbalisable odorants, visual stimuli, and verbal stimuli in respect to PI susceptibility. This proposition is tested directly in Experiment 4 where memory performance in the *n*-back task for visual, verbal, and high/low verbalisable odorants are explored.

Metacognitive judgments further indicate that memory processes for high and low verbalisable odorants may differ. One such metacognitive measure used is the remember/know (K/R) procedure which explores judgments for the type of memory underpinning a retrieval and is premised on the proposition that stimulus recognition comprises both recollection and familiarity judgments (e.g. Wixted & Mickes, 2010; Yonelinas, 2002), and that these processes are dissociated (e.g. Dudukovic, Dubrow, & Wagner, 2009; Evans & Wilding, 2012; Gardiner, Java, & Richardson-Klavehn, 1996; Jacoby, 1991; Koen & Yonelinas, 2016; Olsson et al., 2009; Yonelinas & Jacoby, 1994; for contrast cf. Dunn, 2008). Recollection judgments involve the retrieval of qualitative information about a stimulus (e.g. identification, temporal positioning, spatial location etc., e.g. Gardiner, Ramponi, & Richardson-Klavehn, 1998; Jacoby, 1991; Tulving, 1985; Yonelinas, 1999). In contrast, familiarity judgments involve a graded strength signal (Yonelinas, 1999, 2002) that must exceed a response criterion for correct recognition (Yonelinas, 1999). Previous studies have applied the remember/know procedure to olfactory memory tasks (Larsson, Öberg, & Bäckman, 2006; Olsson et al., 2009) and found verbalisable odorants were associated with ‘remember’ responses to a greater extent than unidentified odorants. Indeed, Olsson et al. argued that recognition for nameable odorants, in comparison to unidentified odorants, more closely resembled words (e.g. remember responses were affected by encoding depth for nameable odorants and words, but not for unidentified odorants).

This purported difference in the type of memory used for high and low verbalisable odorants may explain the differences in *n*-back performance between high and low verbalisable odorants (Jönsson et al., 2011). Recollective judgments in working memory have been linked to controlled working memory resources (Baddeley, 2012; Barrett, Tugade, & Engle, 2004; Loaiza, Rhodes, Camos, & McCabe, 2015); it is therefore possible that a high control strategy, such as rehearsing the last *n* items within memory, may be employed for high verbalisable odorants whereas an alternative strategy is used for low verbalisable odorants. This reliance on different types of memory for high and low verbalisable odorants in the *n*-back task is explored in Experiment 3. Indeed, more generally it is not clear how odorants are maintained in working memory in order to perform the *n*-back task. If one assumes that the odorants (particularly low verbalisable odorants) are not being rehearsed as verbal labels, one candidate mechanism is attentional refreshing. This refers to a domain-general maintenance process that keeps a representation active in attention (see Camos, Johnson, Loaiza, Portrat, Souza, & Vergauwe, 2018, for review). The important point being that this process is non-verbal and has been linked to the maintenance of items within the episodic buffer and visuo-spatial sketchpad (Baddeley, 2012). It has been argued by some (e.g. Loaiza & McCabe, 2012; Oberauer, 2002) that a key feature of refreshing is to strengthen the association that an item has with a context (e.g. binding an item to a temporal context); this makes refreshing a good fit for the *n*-back task as accuracy is dependent upon recollecting the position of the test probe in a preceding sequence. Refreshing is, however, argued to be attentionally demanding (Barrouillet, Bernardin, & Camos, 2004). In Experiment 2 we explore whether olfactory *n*-back performance is affected by a secondary attentionally-demanding task that should disrupt refreshing.

The present set of experiments are designed to both explore in detail the *n*-back performance advantage for verbalisable odorants (Jönsson et al., 2011) and examine the extent to which this quantitative difference reflects the employment of different memory processes. Our overarching question concerns the extent to which working memory (assessed using the *n*-back task) for olfactory stimuli is different to that of other more researched stimulus types (e.g. visual and verbal stimuli) and whether this is moderated by the verbalisability of the odorants.

Experiment 1

Experiment 1 is designed as a partial replication of Jönsson et al. (2011). The purpose of this replication is threefold. First, above chance olfactory *n*-back performance has been shown in only two previous studies (Dade et al., 2001; Jönsson et al., 2011) and shown once with hard-to-name odorants (Jönsson et al., 2011). We seek to replicate this finding with a different set of (high and low verbalisable) odorants. Second, this experiment can validate recent normative data from our laboratory (Moss et al., 2016), by demonstrating the same facilitative effect for highly verbalisable odorants as that reported by Jönsson et al. (2011). Third, an increased number of *n*-back trials were used to investigate discriminability changes due to perceptual and verbal learning throughout the task (Jönsson et al., 2011, Experiment 2, employed 25 critical *n*-back trials with high verbalisable odorants and 25 critical trials with low verbalisable odorants). That is, we examine the extent to which performance improves for low verbalisability odorants as a result of repeated exposure to those odorants (in Experiment 1 we employ 48 critical *n*-back trials with high verbalisable odorants and 48 critical trials with low verbalisable odorants). It is possible that repeated exposure to the odorants enables the development and refinement of verbal labels and this process should manifest in a gradual improvement in memory for the low verbalisable odorants compared to the high verbalisable odorants which already possess a verbal label.

In the present experiments we analyse an index of recognition ability using signal detection (hits and false alarms are also reported). Signal detection is preferred over hits as a dependent measure because it takes into account response biases. This is of particular importance when examining the effects of odorant verbalisability, because unidentified odorants have been shown to elicit a greater proportion of rejections at test (i.e., responding that the test item is novel) compared to identified odorants (Frank et al., 2011). Fewer "old" responses at test (i.e., responding that the test item was presented previously) therefore decrease the likelihood of responding correctly for targets by chance. This response bias will, therefore, artificially deflate the proportion of correct responses for targets ('hits') in the low verbalisable condition. Moreover, *A'* is used as the measure of signal detection due to the unequal number of targets and lures in the procedure (see also Jönsson et al., 2011).

For Experiment 1 three hypotheses are presented based upon previous evidence of an *n*-back working memory advantage for verbalisable odorants. It is predicted that (1) working memory performance will be above chance for low verbalisable odorants but, (2) performance for high verbalisable odorants will be significantly better than that for low verbalisable odorants (Jönsson et al., 2011). Across testing blocks, if repeated exposure to odorants leads to the development and refinement of verbal labels (Nguyen, Ober, & Shenaut, 2012; Stevenson, 2001; Stevenson & Mahmut, 2013a), we predict (3) a gradual performance improvement for low verbalisable odorants compared to high verbalisable odorants which already possess a verbal label.

Method

Participants

Twenty Bournemouth University undergraduates (12 males and 8 females, mean age = 20.0, *SD* = 2.7) participated in exchange for course credit. Participants who self-reported olfactory impairments (e.g. symptoms of cold) and smoking (Katotomichelakis et al., 2007) were excluded, as were participants aged over 40 years (Doty et al., 1984). Ethical approval was obtained via the Bournemouth University Ethics Committee.

Materials

The odorants were selected from a corpus of 200 food and non-food related odorants, prepared by Dale Air Ltd. (www.daleair.com), on the basis of normative scores reported by Moss et al. (2016). Each odorant comprised 5ml of an oil-based liquid stored in an opaque test tube in order to mask the odorant's colour. Twelve odorants were selected for use in the *n*-back task; 6 were selected at random from the 20 highest verbalisability scores and 6 were selected at random from the 20 lowest verbalisability scores (Table 1). The verbalisability scores are provided in Moss et al. (2016), and follow a modified version of the methodology described by Jönsson et al. (2011). In Moss et al. (2016), each odorant was scored from 0-3 according to the quality of the verbal labels provided, with a lower score indicating vague or absent labelling and a higher score reflecting use of a specific noun (although this did not need to be the veridical label). For the present 12-odorant stimulus set, verbalisability scores were significantly higher for the high verbalisable odorants, $t(12) = 26.38$, $p < .0005$, $d = 15.23$, $BF_{10} > 1{,}000$ ($M_{high} = 2.66$, $SD_{high} = 0.11$; $M_{low} = 1.12$, $SD_{low} = 0.09$).

Two additional (non-analysed) odorants were used as buffer items at the start of each testing sub-blocks. One odorant was additionally selected from the high verbalisable odorant set and one was additionally selected from the low verbalisable odorant set.

- Table 1 about here please -

Prior to the olfactory *n*-back task, participants completed a visual version of the task as a practice trial. Eight line drawings, each printed on individual A5 sheets of paper, were taken from Snodgrass and Vanderwart (1980) and used as the 2-back practice task. This practice task comprised presentation of a 10-item sequence with eight lure trials (requiring a ‘no’ response) and two target trials (requiring a ‘yes’ response). The drawings were presented on individual pages in a booklet and the experimenter turned over each page in the booklet to display each drawing. Participants responded verbally ‘yes’ or ‘no’ according to the 2-back task criteria, described below, with each image displayed until participants provided a response. The function was to ensure that participants understood the task demands and no participants required more than 2-repetitions of the practice trial.

Design

A 2-factor (2x2) within-participants design was employed. The first factor refers to odorant verbalisability (high versus low verbalisability) and the second to testing block (first or second). Each block contained 52 *n*-back trials, where a trial comprised the presentation of an odorant followed by a judgment as to whether that odorant matched the odorant presented two trials earlier. The 52 trials within each block were divided further into two sub-blocks: one sub-block of 26 trials comprising high verbalisable odorants and a second sub-block of 26 trials comprising low verbalisable odorants. Presentation order of the sub-blocks was counterbalanced across participants. Presentation of the high and low verbalisable odorants was blocked, rather than mixed, as Jönsson et al. (2011) reported stronger effects using a blocked design.

Each 26 trial sub-block commenced with two presentations of the (non-analysed) buffer odorant, e.g. for the high verbalisable odorant sub-block 'pear' is presented twice at the start. Following double presentation of the buffer odorant, the remaining 24 experimental trials comprised the same 6-odorants each being presented on four occasions: once as a 'target' (25% of trials), and three times as a 'lure'[2] (75% of trials). This, therefore, resulted in 6 target trials and 18 lure trials within each sub-block. Target odorants were always presented two trials earlier, and required a 'yes' response. In contrast, lure odorants were never presented two trials earlier, and required a 'no' response. The positioning of the targets and lures within the sequence was pseudo-randomised with the caveat that accidental target trials were avoided. Given that participants undertook four 26-trial sub-blocks across the experiment, four different 26-trial orders were produced that contained a unique order of target and lure trials. These were presented in a counterbalanced order across participants.

Procedure

The experiment was conducted in a quiet, well-ventilated room with a fan to circulate fresh air. Participants were tested individually and sat opposite the experimenter, separated by a wooden screen with a central fixation cross to prevent visual inspection of the odorants. Prior

---

[2] Note that in the present experiment 'lure' refers to all non-targets (as was the case in Jönsson et al., 2011), rather than just non-targets that are close to a potential target position (as used by Kane et al., 2007; Schmiedek, Li, & Lindenberger, 2009). This issue is discussed and addressed in Experiment 4.

to the olfactory task, participants completed the 10-item visual sequence of the 2-back task in order to familiarise themselves with the procedure.

- Figure 1 about here please -

The 2-back task (Figure 1) comprised the presentation of a continuous sequence of odorants, with each odorant then compared to the odorant presented 2 trials previous. Due to the continuous nature of the task, participants are required to update memory through the sequence in order to make comparisons in subsequent trials. Each trial comprised the presentation of a single odorant under the nose of the participant for 2 seconds. Participants were required to make a verbal ‘yes’ response if the odorant matched the odorant presented two trials earlier and a verbal ‘no’ response if it did not. In order to follow the presentation rate of Jönsson et al. (2011), an 8-second inter-stimulus-interval (ISI) separated odorant presentations. There was a 5-minute rest session between the 52-odorant sequences during which participants were encouraged to drink water. The complete experiment lasted approximately 30 minutes.

Analysis

Data in this and subsequent experiments were analysed using traditional analysis of variance (ANOVA) and planned Bonferroni-corrected comparisons. In addition, Bayesian ANOVA with default priors were also performed using JASP (Love et al., 2015; R. D. Morey & Rouder, 2015; Rouder, Morey, Speckman, & Province, 2012). This is a model-based approach, where models containing main effects and interactions can be compared. This process produces a Bayes Factor value that indicates the ratio of support for one model over another. Typically, the comparison will be between a particular model (for example, a model with both main effects) and the null model, producing a Bayes Factor indicating the level of support for this model. However, Bayes Factors are transitive, and the assessment of models additive, so a model with an interaction term added can also be compared to this main effects model. This produces a Bayes Factor indicating the strength of evidence for an improvement to the model when the interaction term is included.

A typical cut-off regarded as providing substantial support for the alternative hypothesis is a Bayes Factor of 3 and a Bayes Factor below 1/3 provides substantial support for the null hypothesis. A score between 1/3 and 3 indicates insensitivity to either hypothesis. Moreover, a Bayes Factor can be interpreted as anecdotal evidence in support of a hypothesis if the direction of evidence points towards the hypothesis but has not reached the required standard of evidence (i.e. >3 or <1/3).

Bayes Factors are also calculated for paired comparisons, which, using the same cut-offs above, outlines the strength of evidence for or against an alternative hypothesis. For the present set of experiments, we use a default Cauchi prior distribution (Rouder et al., 2012). These are presented with *p* values and *t*-test results where appropriate, in the format BF_{10} when testing the alternative hypothesis against the null, and BF_{01} when testing the null hypothesis against the alternative hypothesis. That is, $BF_{10} < 0.33$ and $BF_{01} > 3$ would indicate identical support for the null hypothesis.

The number of correct target identifications (hits) and incorrect identifications of a lure as a target (false alarms, FA) were recorded and used to compute the proportion of hits to FAs via *A'* (as described by Jönsson et al., 2011). This measure of signal detection theory was selected due to the unequal trial numbers for lures and targets, and allows the number of FAs to exceed that for hits. *A'* was calculated as 0.5 + ((hits – FA) x (1 + hits – FA)) / ((4 x hits) x (1 – FA)) when hits exceeded FA, and as 0.5 – ((FA – hits) x (1 + FA – hits)) / ((4 x FA) x (1 –

hits)) when FA exceeded hits. Unlike *d'* where hit rates of one or FA rates of zero result in an indefinite value, use of *A'* allows these results to remain unadjusted.

The same analysis is conducted across the *A'* sensitivity measure, hits, and false alarms.

Results

A' sensitivity

Figure 2a reports *A'* for the high and low verbalisability groups, across experimental blocks. A 2-factor (2x2) within-participants ANOVA was performed with the factors odorant verbalisability (high vs low odorant verbalisability) and block (first and second). A significant main effect of odorant category was evident, $F(1, 19) = 5.95$, $p = .025$, $\eta_p^2 = .24$, reflecting greater sensitivity for the high verbalisability odorants ($M = .84$, $SEM = .01$) compared to low verbalisability odorants ($M = .79$, $SEM = .02$). The main effect of testing block was non-significant, $F(1, 19) = 0.35$, $p = .560$, $\eta_p^2 = .02$, indicating no overall change in recognition sensitivity between the first and the second block. Of interest in respect to the refinement of verbal labels over time, the interaction between odorant verbalisability and experimental block was significant, $F(1, 19) = 8.32$, $p = .010$, $\eta_p^2 = .31$. Bayesian ANOVA indicated strongest support for a model with main effects and an interaction between verbalisability and block (*BF* = 4.42 vs the null model), preferring this model over a main effects model by a factor of 8.75. The interaction indicates that the memory difference between high and low verbalisable odorants was moderated by testing block. This interaction, reported through both frequentist and Bayesian analyses, is shown in Figure 2a and be can explained descriptively by the advantage for high compared to low verbalisable odorants developing in the second block only.

Indeed, to explore this interaction statistically, follow-on Bonferroni-corrected paired comparisons ($\alpha = .025$) and Bayes Factor analyses were conducted. This analysis revealed no difference between the high verbalisability odorants (*M* = .80, *SD* = .09) and the low verbalisability odorants (*M* = .81 *SD* = .10) in the first testing block, *t*(19) = .41, *p* = .690, *d* = .12, BF_{10} = 0.25. However, the second block saw strong evidence for higher recognition sensitivity for high verbalisability odorants (*M* = .88, *SD* = .07) compared to low verbalisability odorants (*M* = .77, *SD* = .16), *t*(19) = -3.58, *p* = .002, *d* = -.92, BF_{10} = 40.40. That is, an effect of greater recognition sensitivity for highly verbalisable odorants was present only in the second testing block. Indeed, this effect appears to be driven by improvement in recognition sensitivity for the high verbalisable odorants across testing blocks. This is shown by significantly higher recognition sensitivity for the high verbalisable odorants in the second block compared to the first block, *t*(19) = -2.92, *p* = .009, *d* = -.91, BF_{10} = 5.69. Whereas in contrast, there was anecdotal evidence (BF_{10} > 0.33) in support of no difference between testing blocks for the low verbalisable odorants, *t*(19) = 1.27, *p* = .220, *d* = .36, BF_{10} = 0.47.

Using a single sample t-test, *A'* sensitivity scores were also analysed against a chance score of 0.5. There was strong evidence for above chance performance for the low verbalisability odorants in both the first, *t*(19) = 13.60, *p* < .001, *d* = 6.24, BF_{10} > 1,000, and second blocks, *t*(19) = 7.63, *p* < .001, *d* = 3.50, BF_{10} > 1,000.

In summary, these analyses demonstrate that in both testing blocks, sensitivity was above chance for the low verbalisable odorants, illustrating that the 2-back task can be performed successfully with low verbalisable odorants. A recognition advantage for high verbalisable was evident in the second block only, suggesting that the advantage for high verbalisable odorants developed over repeated exposure to the odorants.

The same trends as reported for *A'* are observed for both hits and false alarms (see Figures 2b and 2c, respectively). For brevity, the statistics for these dependent measures are available in the supplementary material.

- Figure 2(a-c) about here please -

Discussion

Experiment 1 has replicated the key findings of Jönsson et al. (2011) by demonstrating (1) above chance performance for low verbalisable odorants despite, (2) superior *n*-back performance for verbalisable odorants. Whilst the present findings replicate Jönsson et al. with a different odorant set, these findings also provide validation for the normative ratings reported by Moss et al. (2016). Specifically, odorants identified as highly verbalisable in Moss et al. show the same *n*-back advantage as the odorants identified as verbalisable from Jönsson et al.'s pilot work.

Superior *n*-back performance for high verbalisable odorants is consistent with a range of previous studies showing a memory benefit for nameable odorants (e.g. Frank, et al., 2011; Jehl et al., 1997; Jönsson et al., 2011; Kärnekull et al., 2015; Lyman and McDaniel, 1986). One explanation for this advantage in Experiment 1 is that memory for high verbalisable odorants can be supported by effective verbal labels, such that these labels are used in a high-control verbal rehearsal strategy (e.g. Chatham et al., 2011; Juvina & Taatgen, 2007). In contrast, low verbalisable odorants cannot effectively employ such a strategy because of impoverished or unrefined labelling.

However, an unexpected finding for Experiment 1 was that the advantage between high and low verbalisable odorants was evident for the second block of trials only. This finding contrasts with that of Jönsson et al. (2011) where superior recognition for high verbalisable odorants was found following 50 trials (the equivalent number of trials to our first block). One possible explanation for this interaction between block number and verbalisability in the present study is that for high verbalisable odorants participants initially ascribed poor quality verbal labels that were gradually refined following multiple stimulus exposures (see Stevenson, 2001). Alternatively, one might argue that over trials, the association between the labels and the odorants strengthens, thereby enabling the label to be used as more effective additional information. This unexpected finding is re-tested in the subsequent experiments.

In summary, Experiment 1 has replicated both the above chance *n*-back performance for low verbalisable odorants and superior sensitivity for high verbalisable odorants reported by Jönsson et al. (2011). Unexpectedly, the advantage for high verbalisable odorants was only found in the second block, whereas performance for low verbalisable odorants did not improve following repeated exposure to the stimuli. The role of experiment block will be examined further in the following experiments but it is also worth future studies considering the extent to which *n*-back peformance and strategies change over multiple blocks.

Experiment 2

The superior *n*-back performance for verbalisable odorants reported in Experiment 1 can be interpreted as reflecting participants' verbal recoding and rehearsal of the high verbalisable odorants. However, we note that the difference between high and low verbalisable odorants reported in Experiment 1 was evident for the second block of trials only. This suggests that verbal recoding and rehearsal of the high verbalisable odorants developed following repeated exposures to those odorants.

In Experiment 2, we test directly the assumption that the *n*-back benefit for verbalisable odorants emerges as a consequence of verbal labelling and rehearsal of those verbal labels. This is achieved via a dual-task paradigm in which a secondary concurrent verbal task is employed. Concurrent articulation (CA) is a secondary task commonly employed for suppression of the articulatory rehearsal process (e.g. Cocchini et al., 2002). This methodology is known to remove the effects of both word length (e.g. Baddeley et al., 1975) and phonological similarity (e.g. Baddeley et al., 1984; Saito, Logie, Morita, & Law, 2008) for visually presented words, suggesting that the conversion of verbal stimuli into phonological representations is disrupted. Here, by parallel reasoning, we aim to disrupt the verbal recoding and rehearsal of odorants using CA. In addition, we include a mental rotation task during the inter-trial interval as a control secondary task (there is some unpublished evidence that mental rotation impairs performance on a verbal *n*-back task due to increased use of executive resources, Simmons, 2000). The mental rotation task allows examination of the extent to which any observed detrimental effects of CA may be explained by the act of performing a secondary task per se, rather than the specific disruption of verbal processing.

Indeed, the inclusion of secondary tasks allows us to directly test the purported role of attentional refreshing in the olfactory *n*-back task. Refreshing is argued to be attentionally demanding (Barrouillet et al., 2004) and has been shown to be disrupted by the inclusion of a secondary task (Barrouillet et al., 2004; Barrouillet, Portrat, & Camos, 2011; Vergauwe, Barrouillet, & Camos, 2010). Given that the attentional demands of refreshing are argued to be domain general (see Camos et al., 2018, for review), it should not matter that the secondary task involves a different modality stimulus (e.g. visual-spatial representations) as attention is drawn from the same finite pool of resources. Moreover, if *n*-back of high verbalisable odorants is more reliant on verbal coding (and therefore verbal rehearsal) compared to the refreshing-reliant low verbalisable odorants, one might predict that high verbalisable odorants would be more resilient to the secondary mental rotation task. This is because sub-vocal rehearsal is argued to be less reliant upon attentional resources compared to refreshing (Camos & Barrouillet, 2014). In contrast, if high verbalisable odorants are reliant upon verbal rehearsal, one might predict that CA would impair *n*-back performance to a greater extent than mental rotation.

Previous studies applying CA to olfactory memory tasks have produced mixed results. For example, Andrade and Donaldson (2007, Experiment 2) found that a secondary digit recall task did not impair performance on a primary olfactory yes/no recognition task. In contrast, Miles and Hodder (2005) reported a main effect of CA on a 2-alterntaive forced choice odorant recognition task (a similar but non-significant trend was also observed in Reed, 2000). The extent to which these findings can be extrapolated to our *n*-back task is questionable, however. Specifically, the memory demands for the *n*-back procedure (continually updating memory and making recognition judgments regarding the item occurring *n* items earlier) differ to those for standard recognition tasks. It is, therefore, possible that these varying task demands differentially affect the requirement for verbal recoding. Additionally, whilst the aforementioned studies (Andrade & Donaldson, 2007; Miles & Hodder, 2005; Reed, 2000) used commercially available aromatherapy odours, the extent to which these odorants were amenable to verbalisation is unknown. This is of importance because we suggest that CA should differentially affect memory for high and low verbalisable odorants.

Experiment 2, thus, tests directly the extent to which the *n*-back working memory benefit for high verbalisable odorants is due to verbal labelling and rehearsal of those verbal labels. Specifically, this is the first experiment to apply CA to the olfactory *n*-back task. We replicate the olfactory *n*-back methodology described for Experiment 1 and include concurrent secondary tasks in the inter-trial interval. It has been suggested that the representation of unfamiliar, hard-to verbalise odorants is determined by the olfactory perceptual code in working memory (e.g. Zelano et al., 2009). In contrast, it is argued that the representation of a verbalisable odorant is determined by both a perceptual code and a verbal label (Paivio, 1990;

Stevenson & Wilson, 2007; Yeshurun, Dudai, & Sobel, 2008). By including CA during the inter-trial interval (a manipulation shown to disrupt verbal rehearsal in a visual *n*-back task, Vuontela, Rämä, Raninen, Aronen, & Carlson, 1999), we disrupt the rehearsal of those odorant-derived verbal labels. We predict, therefore, greater impairment for high verbalisable odorants compared to low verbalisable odorants due to the former's greater reliance on verbal labels. By including the secondary mental rotation task, we test directly the extent to which test performance of a secondary task, irrespective of secondary task modality, impairs olfactory *n*-back performance. This directly tests the proposition that refreshing is the mechanism by which odorants are maintained within working memory. Since refreshing is purportedly attentional demanding and sensitive to secondary tasks (Barrouillet et al., 2004; Barrouillet et al., 2011; Vergauwe et al., 2010), we predict a main effect of secondary task. Moreover, if high verbalisable odorants are supported by verbal rehearsal of labels (a process that is less reliant upon attention, Camos & Barrouillet, 2014), we predict secondary task to interact with verbalisability of odorant, such that low verbalisable odorant memory will be disproportionately impaired due to the disruption of refreshing by increased attentional load.

Additionally, in Experiment 2 we seek to replicate the unexpected finding from Experiment 1, that the recognition advantage for high verbalisable odorants is evident for the second block only.

Method

Participants

Seventy-two Bournemouth University undergraduates (mean age = 19.82, SD = 2.93, 61 females, 11 males) participated in exchange for course credit and were randomly assigned in equal numbers to one of three experimental groups. None had participated in Experiment 1, and the same exclusion criteria as described for Experiment 1 were applied.

Materials

The olfactory stimuli were as described for Experiment 1.

The mental rotation task consisted of 104 pictures with each comprising pairs of horizontal 3-dimensional objects in the style of that reported by Shepard and Metzler (1971) and obtained from Ganis and Kievit (2015). The 3-dimensional objects were white on a black background and comprised 8-11 cubes connected end-to-end to form a continuous shape with four arms (see Figure 3). The left object of the pair functioned as the baseline object with which the right target object of the pair was compared. In 50% of the trials the baseline and target objects were identical (congruent trials). In the remaining trials, the objects differed (incongruent trials) such that the baseline and target objects were identical with the exception of a single arm pointing in the opposite direction for the target object. In each of these trials the target object was a rotation (clockwise on a vertical axis 100 or 150 degrees) of the baseline object, with participants required to mentally rotate the target object to determine whether it matched the baseline object. Images were 800x427px and displayed in the centre of a 22-inch 60Hz monitor using stimulus presentation software OpenSesame (Mathôt, Schreij, & Theeuwes, 2012).

Design

A 3-factor (3x2x2) mixed design was employed. The between-participants factor was secondary task type (no secondary task, CA, and mental rotation). The first within-participants factor was odorant verbalisability (high or low) and the second was experimental block (first or second). The construction of blocks and sub-blocks was as described for Experiment 1. The dependent variables were *A'* sensitivity, hits, and FAs.

Procedure

The *n*-back procedure followed that described for Experiment 1 with the exception that participants made their 2-back decision using a 7-button Cedrus Response Box (Cedrus Corporation, San Pedro, USA). Participants pressed the left button for a 'no' response and the right button for a 'yes' response using the index and middle finger of their dominant hand (speeded responses were not requested nor were reaction times analysed). The primary difference compared to Experiment 1 concerned the between-participants inclusion of concurrent task (see Figure 3).

*Concurrent articulation group***.** Participants (*N* = 24) were required to count repeatedly (1, 2, 3, 4, 1, 2…) throughout the 8-second inter-trial interval. Participants were presented with an odorant and made a 'yes' or 'no' 2-back match response. Participants then immediately counted out loud at a rate of approximately 2-3 digits per second, for 8-seconds until presentation of the next odorant.

Concurrent mental rotation group. Participants (*N* = 24) were required to perform a visual mental rotation task during the 8-second inter-trial interval. Participants received a single pair of adjacently positioned 3-dimensional objects presented simultaneously on the screen and were required to judge whether the two objects were identical. Responses were made on a computer keyboard with stickers over the 'z' key for a 'no' response, and 'v' key for a 'yes' response. Participants made a 'yes' response if the objects were congruent and 'no' response if non-congruent. Following a response, the objects disappeared from the screen.

Within each 26-trial sub-block of the olfactory *n*-back task, participants therefore completed 26 mental rotation trials. Across these sub-blocks there were an equal number of congruent and non-congruent trials. The presentation order of the congruent and non-congruent trials was randomised.

Control group. The control group ($N = 24$) performed the *n*-back task as described for Experiment 1.

- Figure 3 about here please -

Results

Figure 4(a-c) displays recognition sensitivity (*A'*), the proportion of hits, and the proportion of false-alarms across the three experimental groups and is collapsed across block. The experimental block variable is not included because, as can be seen below, there was evidence against any main effect or interaction with block number.

A' sensitivity

Figure 4a shows the mean *A'* for each of the three experimental groups across the two odorant verbalisability conditions, collapsed across block number. A mixed 3-factor (3x2x2) ANOVA was computed with the between-participants factor secondary task type (no task, CA, and mental rotation), and the within-participants factors odorant verbalisability (high versus low) and experimental block (first versus second). The main effect of secondary task type was non-significant, $F(2, 69) = 0.91$, $p = .407$, $\eta_p^2 = .03$. There was a significant main effect of odorant verbalisability, $F(1, 69) = 10.64$, $p = .002$, $\eta_p^2 = .13$, with poorer performance for low verbalisability odorants (*M* = 0.76, *SEM* = 0.01) compared to high (*M* = 0.80, *SEM* = 0.01). The main effect of testing block was non-significant, *F*(1, 69) = 0.37, *p* = .3548, η_p^2 = .01. The predicted 2-way interaction between secondary task type and odorant verbalisability was non-significant, *F*(2, 69) = 0.24, *p* = .785, η_p^2 = .01, as was the predicted 3-way interaction between secondary task type, odorant verbalisability, and experimental block, *F*(1, 69) = 0.27, *p* = .762, η_p^2 = .01. In contrast to Experiment 1, the 2-way interaction between odorant verbalisability and experimental block was non-significant, *F*(1, 69) = 0.27, *p* = .603, η_p^2 < .01.

Bayesian ANOVA indicated strongest support for a model with a single main effect of odorant verbalisability (*BF* = 15.15 vs a null model), and that this model was strongly preferred over all interaction models.

- Figure 4(a-c) about here please -

Hit rates

Figure 4b shows the mean hit rate for each of the three experimental groups and two odorant verbalisability conditions, collapsed across block number. The same ANOVA as described for *A′* was computed and revealed non-significant main effects of secondary task type, *F*(2, 69) = 1.35, *p* = .267, η_p^2 = .04, odorant verbalisability, *F*(1, 69) = 3.11, *p* = .082, η_p^2 = .04, and experimental block, *F*(1, 69) = 0.53, *p* = .467, η_p^2 = .01. The predicted 2-way interaction between secondary task type and odorant verbalisability was non-significant, *F*(2, 69) = 0.34, *p* = .690, η_p^2 = .01, as was the predicted 3-way interaction between secondary task type, odorant verbalisability, and experimental block, *F*(2, 69) = 0.40, *p* = .668, η_p^2 = .01. In contrast to Experiment 1, the 2-way interaction between odorant verbalisability and experimental block was non-significant, *F*(1, 69) = 2.08, $p = .154$, $\eta_p^2 = .03$.

Bayesian ANOVA revealed all models to be in favour of the null. Analysis of the hit rates, therefore, shows no difference between the conditions.

False Alarms

Figure 4c shows the false alarm rates for low and high verbalisability odorants, across the three experimental groups. The same ANOVA as described for *A'* was conducted. Analysis revealed a significant main effect of secondary task type, *F*(2, 46) = 3.92, $p = .024$, $\eta_p^2 = .10$. Post-hoc Bonferroni-corrected comparisons revealed a significantly higher false alarm rate ($p_{bonf} = .030$) in the mental rotation task group (*M* = .26, *SEM* = .02) than the control group (*M* = .20, *SEM* = .02). A non-significant difference was observed ($p_{bonf} = .18$) between false alarms in the mental rotation task group and the CA group (*M* = .22, *SEM* = .02), and a non-significant difference was found between the CA and quiet groups ($p_{bonf} > .999$). A significant main effect of odorant verbalisability was also observed, where there were more false alarms for low verbalisability odorants (*M* = .25, *SEM* = .01) than high verbalisability odorants (*M* = .20, *SEM* = .01), *F*(1, 69) = 18.14, $p < .001$, $\eta_p^2 = .21$. This was inconsistent with Experiment 1, where the main effect of verbalisability was non-significant. A non-significant main effect of experimental block was observed, *F*(1, 69) = 3.02, $p = .087$, $\eta_p^2 = .04$. The predicted 2-way interaction between secondary task type, and odorant verbalisability was non-significant, *F*(2, 69) = 1.42, $p = .250$, $\eta_p^2 = .04$, as was the predicted 3-way interaction between secondary task type, odorant verbalisability, and experimental block, *F*(2, 69) = 0.12, $p = .884$, $\eta_p^2 < .01$. Contrasting Experiment 1, the 2-way interaction between odorant verbalisability and experimental block was non-significant, *F*(1, 69) = 1.02, $p = .316$, $\eta_p^2 = .02$.

Bayesian ANOVA revealed the best model to contain main effects of verbalisability and secondary task group (1010.97 vs. the null model). This model got substantially worse by the inclusion of all main effects, or the inclusion of any interaction terms. Together, these findings suggest evidence against any interaction with concurrent task group or with experimental block.

Discussion

Experiment 2 is the first to apply CA to an olfactory *n*-back task and we report four key findings. First, we replicate the performance advantage for high verbalisable odorants reported for Experiment 1 (see also Jönsson et al., 2011). Second, in Experiment 2 we fail to replicate the unexpected finding from Experiment 1 that the advantage for high verbalisable odorants is found only in the second experimental block. Here we find this advantage in both experimental blocks (consistent with Jönsson et al., 2011, who found the effect in a single set of ~50 trials). It is not clear why the effect of block has not been replicated. The only methodological difference between Experiment 1 and the control condition in Experiment 2 is how responses are outputted (verbally in Experiment 1 and via a response box in Experiment 2). It is possible that the verbal outputting process in Experiment 1 attenuated the development of verbal labelling and therefore delayed the onset of the verbalisability advantage. However, this account is speculative and requires further investigation. Third, we report no main effect of secondary task (with the exception of FAs, see below) and no interaction between secondary task and verbalisability of the odorants. Specifically, CA did not impair *n*-back performance overall or disproportionately affect memory for high verbalisable odorants. This finding is important given our working hypothesis that verbal rehearsal of the odorants' names underpins the memory advantage for high verbalisable odorants. However, despite restricting verbal rehearsal opportunities (via CA), the advantage for high verbalisable odorants persisted.

Fourth, related to the preceding point, the evidence for a secondary task disrupting overall *n*-back performance was equivocal. Whilst there was no main effect of secondary task on *A'* and hits, FAs were significantly increased following mental rotation relative to the control condition. This provides only partial support for the proposition that refreshing is used during the olfactory *n*-back task, as the secondary task only affected FAs.

That mental rotation nominally affected olfactory *n*-back performance is consistent with refreshing being employed to maintain odorants within working memory. Given that refreshing is attentionally demanding (Barrouillet et al., 2004), the secondary mental rotation task used finite domain general attentional resources (Camos et al., 2018) that would otherwise be employed in refreshing. However, there are two important caveats to this observation. First, the main effect of secondary task was only found with FAs and was absent for both *A'* and hits. This suggests that loading attention (via mental rotation) had neither a generalised nor catastrophic effect on task performance. Indeed, overall memory sensitivity (*A'*) was unaffected by secondary task. One possible explanation is that mental rotation was insufficiently taxing to induce large effects, with FAs more sensitive to load (e.g. see Pelegrina et al., 2015, for evidence for hits and FAs being differentially affected by age). A second caveat is that the effect of secondary task on FAs was found only following mental rotation and not CA. It may be the case that CA employs limited attentional resources (see Camos & Barrouillet, 2014) and therefore employs only limited resources from the domain general pool. However, an alternative explanation for the detrimental effect of mental rotation was highlighted during the review process. Experiment 2 changed the response format (and modality of response) from a verbal output to a button press. It is possible that use of the response box added an additional spatial component to the task and may have contributed to the increase in false alarms following mental rotation. Similarly, both the primary *n*-back task and the secondary mental rotation task required yes/no responses (albeit one via a response box and one via a keyboard press). It is also possible that interference between these identical response judgments caused the impairment. Notwithstanding these alternative explanations, it is important to emphasise that the detrimental effects of secondary task on olfactory *n*-back performance were limited.

That CA did not impair olfactory *n*-back performance is consistent with the proposal that olfactory working memory is supported by an olfactory, rather than verbal, slave system (Andrade & Donaldson, 2007). If one accepts that CA disrupts the sub-vocal rehearsal of verbal labels (although cf. Jalbert, Neath, & Surprenant, 2011), Experiment 2 suggests that verbal rehearsal for the high verbalisable odorants cannot account solely for the memory advantage for these odorants. If verbal rehearsal is unable to explain the *n*-back advantage for verbalisable odorants, then what does enable superior recognition? One possible candidate mechanism is highlighted by Jönsson et al. (2011) who reported that the high verblisable odorants in their study were easier to discriminate (tested by a same/different judgment on odorant pairs). If one can more accurately distinguish between odorants at the perceptual stage, it follows that there will be fewer errors at encoding and, therefore, fewer subsequent errors at the recognition phase. This perceptual advantage for the high verbalisable odorants may derive from initial verbal recoding during the presentation phase of the odorant (i.e. whilst sniffing an odorant, using a label accentuates the differences between items). Alternatively, the memory advantage for the verbalisable odorants might be explained by the high verbalisable odorants having greater familiarity for the participant (Moss et al., 2016, report a strong positive correlation between verbalisability and familiarity: $r = .88$). Indeed, it has been proposed that exposure to familiar odorants activates more specific olfactory representations in memory compared to unfamiliar odorants (Stevenson & Mahmut, 2013b; Wilson & Stevenson, 2006). It follows that more specificity in the olfactory representation would lead to less confusion in memory judgments due to high verbalisable odorants being more perceptually discrete. Indeed, if familiar odorants are easier to differentiate at both the encoding and test phases (e.g. see also

Rabin, 1988), then fewer working memory resources will be required when performing the task (Reder, Liu, Keinath, & Popov, 2015). That is, fewer resources are needed to resolve the perceptual confusions between odorants. Given that the type of representation activated for high and low verbalisable odorants has been argued to differ (Stevenson & Mahmut, 2013b; Wilson & Stevenson, 2006), we suggest that participants employ different processes in performing the *n*-back task for high and low verbalisable odorants. Experiment 3 is designed to explore evidence for the proposition that qualitative differences exist between the *n*-back memory processes used for high and low verbalisable odorants by including metacognitive judgments after each *n*-back response.

Experiment 3

Experiment 2 replicated the *n*-back performance advantage for high verbalisable odorants. Our working hypothesis was that this advantage is a result of verbal labelling and rehearsal of the high verbalisable odorant. However, in contrast to our prediction, this advantage was not attenuated by a secondary concurrent verbal task. Our finding thereby contradicts verbal rehearsal during the inter-trial-interval as a viable mechanism underpinning the *n*-back performance difference between the high and low verbalisable odorants. Experiment 3 is designed to examine the extent to which the performance difference can be explained by different memory processes being used for high and low verbalisable odorants. Our rationale for this investigation is premised on the finding that the metacognitive judgments applied to identified and unidentified odorants differed (Larsson et al., 2006; Olsson et al., 2009). Specifically, they found that identified/nameable odorants exhibited more recollection judgments compared to unidentified odorants. Experiment 3 uses the remember/know procedure to identify if different types of memory processes (e.g. familiarity versus recollection) support the *n*-back task for high and low verbalisable odorants.

The application of the remember-know procedure to the *n*-back task is quite different to its conventional use in yes/no recognition tasks. In these tasks (e.g., Evans & Wilding, 2012; Koen & Yonelinas, 2016; Tulving, 1985; Yonelinas, 2002), following presentation of a long sequence, participants are required to provide a binary 'old' or 'new' recognition judgment for a test item (with K/R judgments following an 'old' response). That is, participants judge whether the item is experimentally novel. However, in the *n*-back task the memory demand is different, with participants required to judge whether the current item is the same item as that presented 2-items earlier. Indeed, given that the same 6 odorants are used throughout each presentation block, odorants rapidly cease to be experimentally novel. Therefore, in our *n*-back task, and in contrast to the yes/no recognition task, participants are not deciding whether the probe item is experimentally novel. As a result, the *n*-back metacognitive judgement requires participants to judge not only whether a test odorant is 'old' but, in addition, to judge 'how old'. To be clear, in the *n*-back task it is not an effective strategy to use familiarity dichotomously (i.e., familiar or unfamiliar), since all items will be experimentally familiar. Instead, for an effective familiarity (*K*) response, participants must assess how long ago an item was presented based upon the strength of familiarity for the probe item (e.g. see Schmiedek, Li et al., 2009; Oberauer, 2005). Specifically, the strength of the familiarity signal for the probe item could be used to determine whether the probe was presented 2-trials previous (e.g., the item is very familiar suggesting it was previously presently very recently and potentially warrants a "yes" response). In contrast to familiarity, an *n*-back recollection (*R*) judgment requires positional recall, presumably as a result of a successful binding between item and context (Oberauer, 2005). That is, participants explicitly recall that this test probe was presented two trials previous. Such a judgment is consistent with participants storing the last *n*-items within a dynamically updated maintenance window. In addition, the present experiment also employs a guess (*G*) response as one of the metacognitive judgments. A *G* response is used in the *n*-back task in the same way that it is employed for other recognition tasks; that is, for situations in which a correct response is made in the absence of any recollective experience or strategy (Gardiner, Ramponi, & Richardson-Klavehn, 2002).

Experiment 3 is designed to replicate the methodology of Experiment 1, with the additional requirement for a metacognitive judgement. Specifically, after each ‘yes’ response, participants are required to provide a *K* (know, i.e. familiarity), *R* (remember, i.e. recollection), or *G* (guess) judgment. The inclusion of such metacognitive judgments is to determine differences in *n*-back strategy between high and low verbalisable odorants. We argue that different memory processes (e.g. using familiarity, recollection etc.) may indicate the employment of different *n*-back strategies. This claim is based upon recollective judgments being linked to controlled working memory resources (Baddeley, 2012; Barrett et al., 2004; Loaiza et al., 2015), e.g., active rehearsal/refreshing of the last *n* items. We therefore suggest that *R* responses are indicative of participants actively rehearsing and updating items (akin to the conventional view of how working memory is employed in the *n*-back task, e.g. Kane et al., 2007). Here, we directly compare whether the proportion of *R* responses differs between high and low verbalisable odorants. Indeed, given that high, compared to low, verbalisable odorants are more likely to be recollected (Larsson et al., 2006; Olsson et al., 2009), it is predicted that high verbalisable odorants will exhibit a greater proportion of recollection (*R*) responses in the olfactory *n*-back task.

Method

Participants

Twenty-four female Bournemouth University undergraduates (mean age = 20.21, *SD* = 3.19) participated in exchange for course credit. The same exclusion criteria as described for Experiments 1 and 2 were applied and none had participated in Experiments 1 or 2.

Materials

The olfactory stimuli were taken from the same corpus of odorants described for Experiment 1. A set of twelve odorants (6 categorised as high verbalisable and 6 categorised as low verbalisable, and not employed in Experiment 1 and 2, see Table 2) were selected. The high verbalisable odorants ($M = 2.64$, $SD = 0.11$) exhibited significantly higher verbalisability scores than the low verbalisable odorants ($M = 1.34$, $SD = 0.26$), $t(10) = 11.54$, $p < .001$, $d = 6.66$, $BF_{10} > 100$. As noted above, verbalisability correlates strongly with familiarity (Moss et al., 2016) and consequently the high verbalisable odorants also exhibited significantly higher familiarity ratings ($M = 5.83$, $SD = 0.18$) compared to the low verbalisability odours ($M = 3.33$, $SD = 0.25$), $t(10) = 19.87$, $p < .001$, $d = 11.47$, $BF_{10} > 100$.

In Experiment 3, we matched the two groups of odorants on other dimensions assessed in Moss et al. (2016). The high and low odorants did not differ significantly on measures of intensity, $t(10) = 1.81$, $p = .101$, $d = 1.04$, $BF_{01} = 0.83$, or hedonic strength (a measure of each pleasantness rating's deviation from a neutral midpoint), $t(10) = 0.79$, $p = .449$, $d = 0.46$, $BF_{01} = 1.76$.

- Table 2 about here please -

Design

A 2-factor (2x2) within-participants design was employed with the factors odorant verbalisability (high or low) and experimental block (first or second). Unlike Experiment 1, where high and low verbalisable odorant trials were grouped together, high and low verbalisable trials were randomly ordered within a single block. High and low verbalisable trials were presented in a randomised order (i.e. a mixed design) to promote variance in metacognitive responses (i.e. our working hypothesis is that there will be more *R* responses for high verbalisable odorants). That is, with a continuous block of high verbalisable odorants, participants may become self-conscious of repeatedly responding '*R*' and therefore alternate their responses. If the memory processes for high and low verbalisable odorants do indeed differ, using a mixed design will make these differences salient to the participant. We argue that employment of a mixed, rather than blocked, design should not remove the *n*-back advantage for high verbalisable odorants, as although Jönsson et al. (2011) reported a stronger effect for a blocked design (Experiment 2), it was still present following a mixed design (Experiment 1).

In Experiment 3, the number of trials within each experimental block was reduced in order to match exactly the procedure reported in Jönsson et al. (2011, Experiment 1). An experimental block began with 2 non-analysed buffer trials and was then followed by 36 trials (comprising 18 high verbalisable odorants and 18 low verbalisable odorants). To further match Jönsson et al. (2011, Experiment 1), each odorant was presented on three occasions: once as a target (33.3% of trials) and twice as a lure (66.6% of trials). These ratios differ to that used in

Experiment 1. Whilst changing these ratios may affect hits and false alarms (due to the proportion of hits that would be correct following a guessed response), any effects of response bias are accounted for by the use of signal detection (*A'*).

The dependent variables were again *A'* sensitivity, hits, and FAs. In addition, for the hits and FAs, we report the proportion of 'R', 'K', and 'G' responses.

Procedure

The procedure followed that described for Experiment 1 with an additional requirement at the testing stage. Following a 'yes' response, participants were required to provide a metacognitive judgement. Instructions for this response were a modification of those described by Rajaram (1993): an *R* response was required when participants explicitly recollected the odorant and its occurrence in its correct *n*-back position; a *K* response was required when the 'yes' response was based on familiarity for the odorant; and a *G* response was required when participants made a 'yes' decision based on some other reasoning, strategy, or if they were unsure why they had responded 'yes'. As in Experiment 2, responses were made on a Cedrus Response Box and the input recorded using Superlab 4.5 (Cedrus Corporation, San Pedro, USA).

Results

Figure 5(a-c) displays recognition sensitivity (*A'*), the proportion of hits, and the proportion of false alarms as a function of verbalisability. For the latter two analyses, hit and false alarm *guess* responses were removed (see Olsson et al., 2009, for a similar application of this method; although it should be noted that including *guess* responses did not change the outcome from the analysis detailed below). Consistent with Experiment 2, initial analyses revealed null effects of experimental block (all *F*s<1); consequently for the purpose of analysis performance was collapsed across experimental block.

Working Memory Performance

For the *A'* sensitivity, hits, and false alarms paired *t*-tests were conducted comparing the high and low verbalisable odorants. *A'* sensitivity for high verbalisable odorants (*M* = .88, *SD* = .08) was significantly higher than that for low verbalisable odorants (*M* = .80, *SD* = .16), *t*(23) = 3.02, *p* = .006, *d* = .62, BF_{10} = 7.41. Similarly, hits for high verbalisable odorants (*M* = .74, *SD* = .16) were significantly higher than those for low verbalisable odorants (*M* = .58, *SD* = .23), *t*(23) = 3.69, *p* = .001, *d* = 0.75, BF_{10} = 29.70. There was a non-significant difference between high and low verbalisable odorants for FAs, *t*(23) = 1.15, *p* = .264, *d* = 0.23, BF_{10} = 0.39, with anecdotal support for the null hypothesis.

- Figure 5(a-c) about here please -

Metacognitive Responses: Analytical Approach

We analysed all 'yes' responses and divided them into hits (correct responses) and FAs (incorrect responses). We then calculated the proportion of 'yes' responses that were *remember (R), know (K),* and *guess (G)* judgments. This was calculated for both hits and FAs. This is a *relative* calculation that gives the proportion of a response type without consideration to the absolute number of ‘yes’ responses given (see Larsson et al., 2006, for an example of this analysis applied to remember-know responses).[3]

For both hits and false alarms, the proportion of metacognitive responses are analysed using a 2-factor (2x3) within-participants ANOVA, with the factors odorant verbalisability (low versus high) and metacognitive judgment type (*R*, *K*, and *G*).

Metacognitive Responses: Hits

Figure 6a shows the proportion of response types for correct target (‘yes’) responses. With respect to the ANOVA, the main effect of odorant verbalisability was not assessed because the sum proportion of K, R, and G responses totalled 1 for both high and low verbalisable odorants. The main effect of response type was non-significant, $F(2, 46) = 3.19$, $p = .050$, $\eta_p^2 = .12$. However, the theoretically important interaction between verbalisability and metacognitive judgment was significant, $F(2, 46) = 6.48$, $p = .003$, $\eta_p^2 = .22$. A Bayesian ANOVA indicated strong support for a model that included a metacognitive judgment main effect and an interaction between verbalisability and metacognitive judgment ($BF = 350.43$ vs a null model). This model was preferred to a metacognitive judgment main effect model by a factor of 79.42. That is, there is strong evidence for an interaction between experimental condition and metacognitive judgment.

In order to examine this interaction in more detail, the difference between the proportion of responses for low and high verbalisability odorants was compared independently for each response type. Paired *t*-tests revealed fewer *G* responses for high verbalisability odorants (*M* =

---

[3] We computed an alternative analysis of recollective experience that takes into account the absolute number of ‘yes’ responses and this revealed the same pattern of results (applied in Olsson et al., 2009). These results are not reported.

.19, *SD* = .20) compared to low verbalisability odorants (*M* = .39, *SD* = .22), $t(23) = 5.15$, $p < .001$, $d = 1.05$, $BF_{10} = 756.10$. In contrast, there was evidence against a difference in *K* responses between high and low verbalisability odorants, $t(23) = -0.37$, $p = .713$, $d = 0.08$, $BF_{10} = 0.23$. Finally, significantly more *R* responses were found for high verbalisability odorants (*M* = .51, *SD* = .24) compared to low verbalisability odorants (*M* = .33, *SD* = .25), $t(23) = 2.51$, $p = .020$, $d = 0.51$, $BF_{10} = 2.75$ (although the Bayes Factor was anecdotal).

Metacognitive Responses: False Alarms

Figure 6b shows the proportion of response types for false alarms. The ANOVA revealed a main effect of metacognitive judgment, $F(2, 46) = 3.65$, $p = .034$, $\eta_p^2 = .137$, with more *G* responses (see Figure 5b) reflecting less certainty in these erroneous responses. Metacognitive judgment did not, however, interact with odorant verbalisability, $F(2, 46) = 0.36$, $p = .703$, $\eta_p^2 = .015$. The preferred model contained only a main effect of response type (BF = 37.41 vs. a null model).

- Figure 6(a-b) about here please -

Discussion

Experiment 3 is the first to apply the K/R procedure to the olfactory *n*-back task. Consistent with Experiments 1 and 2, we again report an *n*-back advantage for verbalisable odorants (and consistent with Experiment 2, an effect that was found across experimental blocks) and replicate the effect using a different set of odorants. In addition, this effect was evident in Experiment 3 when high and low verbalisable odorants were presented using a mixed trial design (as used in Experiment 1 of Jönsson et al., 2011). Importantly, the inclusion of *know*/*remember*/*guess* judgments revealed different patterns of metacognitive judgments for both the accuracy of response (hit and FA) and the odorant type (high and low verbalisable). A greater proportion of 'guess' responses was observed for false alarms relative to hits (see Figure 5); this reflects less certainty with respect to erroneous responses. Second, a different pattern of metacognitive responses was found for high and low verbalisable odorants. The proportion of hits that received a 'remember' (*R*) response was higher for verbalisable odorants compared to low verbalisable odorants. This finding is, therefore, consistent with greater recollection of contextual details (e.g. temporal position etc.) for high verbalisable odorants. There was no difference in 'know' (*K*) responses between high and low verbalisable odorants, but there were more guess (*G*) responses for low relative to high verbalisable odorants. This latter finding demonstrates a general reduction in certainty for the low verbalisable odorants, suggesting a weaker memory trace for these odorants (see Wixted & Mickes, 2010).

That high verbalisable odorants exhibited more recollection judgments than low verbalisable odorants, is consistent with other olfactory studies that used the K/R procedure (albeit with a different memory task) (Larsson et al., 2006; Olsson et al., 2009). The greater proportion of *R* responses could suggest different strategies being used for high and low verbalisable odorants; with *R* responses indicating that participants were able to remember the position in which the probe was originally presented (i.e. 2 trials previous). Indeed, the ability to remember the position of the test item in the preceding sequence requires maintenance and updating within memory. In contrast, a greater proportion of *guess* responses were reported for the low verbalisable odorants. This reflects less information being accessible for these low verbalisable odorants (see Stevenson & Mahmut, 2013b; Wilson & Stevenson, 2006). However, whilst the proportion of the recollection (*R*) responses was less for low verbalisable odorants, some were reported (33% of all correct responses). This suggests that participants are capable of recollecting low verbalisable odorants in the *n*-back task but these odorants are less amenable to recollection than high verbalisable odorants.

An alternative explanation for the difference in the proportion of *R* and *G* metacognitive judgments for high and low verbalisable odorants concerns the strength of the memory signal. Rather than via the use of different memory processes, the inflated guess rate for low verbalisable odorants may simply reflect a weaker memory trace and reduced confidence for these items. Indeed, that the memory trace is weaker for low verbalisable odorants is supported by consistently lower working memory scores relative to high verbalisable odorants (Experiments 1-3). However, rather than memory for high and low odorants existing along the same continuum, it is worth noting that previous work identified differences in susceptibility to PI between these two types of odorants (Moss et al., 2018); a finding that suggests qualitative differences between these two stimulus types.

It is worthy of note that for both high and low verbalisable odorants, a relatively high proportion of correct responses were based upon familiarity (*K*) judgments (for high and low verbalisable odorants = .302 and .275, respectively). One possible explanation for the use of familiarity in order to perform the task relates to the lack of control over 'lures'. In Experiments 1-3, we made no distinction between 'recent' and 'non-recent' lures. If the lures are experimentally non-recent, it follows that they will possess lower levels of familiarity. In contrast, for hits, the odorant was only presented 2-trials previous and would therefore possess higher levels of experimental familiarity. It is possible, therefore, that participants could be making judgments based upon the strength of the familiarity signal rather than via the use of working memory. This limitation is addressed in Experiment 4.

In summary, Experiment 3 has shown differences in the metacognitive judgments for high and low verbalisable odorants during the *n*-back task. That is, whilst a quantitative difference exists with respect to performance levels, even when only correct responses are analysed, recollection is greater for high verbalisable odorants. Our findings in Experiments 1-3, along with other studies (e.g. proactive interference in Moss et al., 2018; and metacognition in Olsson et al., 2009), suggest that memories for low verbalisable odorants may be qualitatively different to those of other stimulus types. Specifically, our working hypothesis is that memory processes for low verbalisable odorants are different to that of high verbalisable odorants; with high verbalisable odorant memory processes being more similar to that of verbal/visual memory (see also Moss et al., 2018; Olsson et al., 2009; Zelano et al., 2009). In Experiment 4 we test this proposition by examining shared variance for the *n*-back task across different stimulus types. We argue that the employment of similar memory processes will be demonstrated by cross-modal correlations in performance. That is, *n*-back performance for verbal and visual stimuli should correlate with high verbalisable odorants but not with low verbalisable odorants.

Experiment 4

Cross-Modal N-Back Correlations

Taken together, the results of Experiments 1-3 demonstrate that quantitative differences exist in memory for high and low odorants using the *n*-back task. Previous work has suggested that high verbalisable odorants are similar to verbal and visual stimuli with respect to both metacognitive judgments (Larsson et al., 2006; Olsson et al., 2009) and susceptibility to proactive interference (Moss et al., 2018); whereas memory for low verbalisable odorants differs qualitatively to these stimulus types. If the processes underpinning memory for high verbalisable odorants are those underpinning visual/verbal memory, then it follows that performance patterns should positively correlate. Specifically, when we compare *n*-back performance for high verbalisable odorants, visual stimuli, and verbal stimuli, then shared variance should be evident due to common processes governing task performance. In contrast, if memory for low verbalisable odorants is qualitatively different, one might predict an absence of shared variance with verbal and visual *n*-back performance. This prediction is tested directly in Experiment 4 by examining the correlations in *n*-back performance between high verbalisable odorants, low verbalisable odorants, letters, and abstract shapes.

Defining Lure Trials

In Experiment 4 we also include an important methodological amendment to the *n*-back procedure described for Experiments 1-3. As noted above, in our preceding experiments (as in Jönsson et al., 2011) we made no distinction between 'recent' and 'non-recent' lures when reporting false alarms and computing *A'* sensitivity. A recent lure is defined as a trial in which the current test odorant was previously presented in a serial position close to that of the target 2-back serial position, i.e., it was presented 1- or 3-trials earlier. A non-recent lure is defined as a trial in which the current test odorant was previously presented in a trial distant to the target 2-back serial position, e.g. 6-back. It is plausible that participants will adopt different strategies when responding to recent and non-recent lures. For example, for non-recent lures, experimental familiarity for that lure will be relatively low and a participant can use that low familiarity signal to judge that the lure was not presented 2-trials earlier and, therefore, reject the item (a 'no' response). However, if the item was presented 1- or 3-back (a recent lure), when, in relative terms, experimental familiarity for that item is high, then judgements premised on item familiarity will be less effective when differentiating between targets and lures. Because Experiments 1-3 did not distinguish between recent and non-recent lures, it is plausible that, for non-recent lures, participants made their judgements based on the familiarity strength of the test probe. To be clear, it is therefore possible that above chance performance in Experiments 1-3 (and Jönsson et al., 2011) was not a result of using working memory but instead by using the familiarity strength of the probe. Therefore, to provide a more valid measure of olfactory working memory, the present experiment only analyses recent lures to compute *A'* and report false alarms. Specifically, we analyse only those lures that were presented either 1- or 3-back, and, as a consequence, their familiarity closely matches that of the 2-back item.

Predictions

The aim of Experiment 4 is to assess the *n*-back performance correlations between high verbalisable odorants, low verbalisable odorants, consonants (verbal memory), and polygons (non-verbal visual memory). To date, research suggests that memory for verbalisable odours may share memory processes with those underpinning verbal memory (e.g. Larsson et al., 2006; Moss et al., 2018; Olsson et al., 2009; for imaging evidence see Zelano et al., 2009). On this basis, we predict a significant positive correlation between *n*-back performance for high verbalisable odorants and consonants (verbal working memory). Given the reported significant correlation between *n*-back performance for verbal and visual stimuli (r = .66, Schmiedek et al., 2014), and by implication the shared resources used for the tasks (supported by similar brain regions being implicated in *n*-back performance cross-modally, Nystrom et al., 2000; Owen et al., 2005), we additionally predict that *n*-back performance for visual stimuli (polygons) and high verbalisable odorants should also correlate positively. Of particular theoretical interest is the correlation between low verbalisable odorants and both verbal (consonants) and visual (polygons) *n*-back performance. Given that the memory processes for low verbalisable odorants appear to differ to that of visual and verbal stimuli, we predict that the correlation between low verbalisable *n*-back performance and that for visual and verbal stimuli will be weaker.

Method

Participants

Fifty-six Bournemouth University undergraduates (44 females, 12 males, mean age = 23.91, *SD* = 6.64) participated for a course credit requirement. The same exclusion criteria were applied as described for Experiment 1 and none had participated in Experiments 1-3.

Materials

The olfactory stimuli were taken from the same corpus of odorants as described for Experiment 1. A set of twelve odorants (6 categorised as high verbalisable and 6 categorised as low verbalisable, plus two buffer odorants, see Table 3) were selected. The high verbalisable odorants ($M = 2.63$, $SD = 0.11$) exhibited significantly higher verbalisability scores ($M = 1.05$, $SD = 0.24$), $t(12) = 12.96$, $p < .001$, $d = 6.93$, $BF_{10} > 1{,}000$, and significantly higher familiarity ratings ($M = 5.83$, $SD = 0.18$) compared to the low verbalisability odorants ($M = 3.33$, $SD = 0.25$), $t(12) = 22.22$, $p < .001$, $d = 11.88$, $BF_{10} > 1{,}000$. The odour sets did not differ significantly on intensity, $t(12) = -0.19$, $p = .851$, $BF_{01} = 2.21$, or hedonic strength (deviance from a neutral midpoint on the pleasantness rating scale), $t(12) = 0.67$, $p = .515$, $BF_{01} = 1.93$. Responses were collected using a Cedrus Response Box, and recorded using Superlab 5 (Cedrus Corporation, San Pedro, USA).

- Table 3 about here please -

Visual stimuli. Seven irregularly-shaped polygons (taken from Chuah, Maybery, & Fox, 2004, see Figure 7 for example) designed to prevent verbal rehearsal strategies (Attneave, Arnoult, & Attneave, 1956; Smith et al., 1995) were used in the visual 2-back task. These were presented using OpenSesame (Mathôt et al., 2012) in the centre of a 22-inch 60hz monitor as black line drawings on a white background, within a black border square of 62px by 62px (approximately 17mm x 17mm).

Verbal stimuli. Eight phonologically dissimilar consonants were selected (B, F, H, K, M, Q, R, X) (Kane et al., 2007), and displayed individually and centrally on a 22-inch 60hz monitor in size 21pt. monospaced font. Consonants were presented in lower or upper case in order to both minimise visual similarities between the letters and to promote the use of verbal rather than visual memory. Stimulus presentation timings and trial responses (see Figure 7) were controlled by OpenSesame.

Design

A correlational design was employed that compared performance on the high verbalisable odorant 2-back task, the low verbalisable odorant 2-back task, the visual 2-back task, and the verbal 3-back task.

Olfactory 2-back task. The task was a shortened version of that described for Experiment 1. A single 52-trial block was employed using an equal number of high and low verbalisable odorants. As in Experiment 1, the high and low verbalisable odorants were tested in 26-trial sub-blocks (each beginning with 2 buffer odorants), and the order of these sub-blocks was counterbalanced across participants. Excluding the initial 2-buffer odorants, each sub-blocks contained 6 targets and 18 lures. Two different 26-trial sequences were created and they were also counterbalanced across the high and low verbalisable odorants.

An important deviation from Experiment 1 is the focus upon *recent-lures*. Recent-lures are trials in which the test odorant does not match the odorant presented 2-trials earlier (thereby requiring a ‘no’ response) but it does match the odorant presented either 1- or 3-trials earlier. Consequently, a recent-lure will exhibit high levels of familiarity, and, therefore, limit the extent to which participants can use a ‘strength of familiarity’ judgement in order to distinguish between targets and lures (Ralph, 2014). This is because the 1-, 2-, and 3-back items should have similar levels of experimental familiarity compared to *non-recent lures* (e.g. if the odorant was presented 6-items previous). By only using recent lures, the familiarity signal has less diagnostic power in distinguishing between hits and lures. As a result, the task has increased validity as a measure of working memory (i.e. maintaining and updating memory for odorants). Within each block of high and low verbalisable odorants there are 5-6 recent-lures and 6 targets. Reponses for non-recent-lures are now discarded, with the recent-lures used to compute both working memory sensitivity (A’) and false alarm rates.

Visual 2-back task. The visual 2-back task consisted of 2 blocks each of 26 items. Two identical buffer images preceded the 24 critical 2-back trials in each block, and did not occur again in either block. *Recent-lures* were included as described for the olfactory *n*-back task. The presentation order of the two visual blocks was randomised across participants. *A’* was used as the dependent variable.

Verbal 3-back task. Pilot work suggested that performance on the 2-back task with verbal stimuli was close to ceiling (≈ 98%). Consequently, a 3-back task was employed requiring participants to state whether the current item matched the item presented 3 items earlier. Participants completed 3-blocks of 40 trials comprising 32 lures and 8 targets. The lures for each block included 7-10 recent-lures (i.e. the lure had been presented 1-, 2-, or 4-items previous).

- Figure 7 about here please -

Procedure.

Participants were tested individually in a well-ventilated laboratory at Bournemouth University, and each completed the three versions of the *n*-back task in a single session lasting ~1hr. Participants performed the computer-based (verbal and visual) and olfactory *n*-back tasks in a partially-counterbalanced order. Specifically, participants completed either the olfactory or computer-based tasks first, but always completed the two computer-based tasks adjacently (though these were themselves performed in a counterbalanced order).

Olfactory 2-back task. Participants sat opposite the experimenter, separated by a wooden screen. The procedure followed that described for Experiment 1.

Computer-based tasks. The computer-based verbal and visual *n*-back tasks were performed with participants sat approximately 50cm in front of a computer monitor. Instructions were presented to participants prior to each *n*-back task (i.e., the 2-back visual and 3-back verbal task). Participants completed 10 practice trials prior to each version of the *n*-back task.

In the visual 2-back task, each abstract polygon was displayed for 2000ms, followed by a fixation cross presented for 2000ms. Presentation time of the visual stimuli was extended, relative to verbal stimuli (see below), due to pilot data showing that 500ms presentation rates resulted in low performance levels. For targets, participants were required to press the '1' key and for lures participants pressed the ‘3’ key. Responses could be made whilst the test stimulus was presented or during the inter-stimulus-interval.

In the verbal 3-back task, each letter was displayed for 500ms, followed by a fixation cross presented for 2000ms. These timings are based on that described by Jaeggi et al. (2010) for a similar 3-back verbal task. The response procedure followed that described for the visual task.

Results

A different approach to analysis described for previous experiments was undertaken in Experiment 4. Specifically, only recent-lures were used to assess FAs, with non-recent lures discarded. The unique use of recent-lures in calculating the false alarm rate produces an A’ score that is not inflated by easy non-recent lure rejections; this provides a more sensitive assessment of differences in working memory ability, but necessarily at the expense of using fewer trials (Ralph, 2014).

Working Memory (N-Back) Performance

A’ sensitivity. Across the four *n*-back tasks (high verbalisable odorants, low verbalisable, letters, and polygons) above-chance performance was assessed using one sample t-tests against an *A’* score of 0.5. All comparisons were significant (*ps.* < .001), and the data strongly supported above-chance performance, $BF_{10} > 1{,}000$.

Figure 8 shows the mean *A’* sensitivity score across the four *n*-back tasks. Any direct comparison between performance levels of the different modality *n*-backs is confounded by methodological differences (e.g. presentation times, size of *n* etc.); therefore only direct comparison between *A’* scores for the high and low verbalisable odorants was conducted. Consistent with Experiments 1-3, *A’* scores were significantly superior for the high, relative to the low, verbalisable odorants, $t(55) = 2.930$, $p = .005$, $d = 0.392$, $BF_{10} = 6.656$.). This finding is of importance, because it replicates the verbalisable advantage reported in Experiments 1-3 using a more stringent calculation of *A'* sensitivity.

- Figure 8 about here please -

Correlational Matrix

Table 4 shows a correlation matrix computed for *A'* scores as a function of task type. Comparisons between *n*-back performance revealed anecdotal ($BF < 3$) support for a moderate positive correlation between low and high verbalisable odorant working memory, $r = .27$ (95% CI [.005,.495]), $p = .047$, $BF_{10} = 2.24$. For high verbalisable odorants there was support for moderate correlations with both and visual, $r = .30$ (95% CI [.035,.518]), $p = .027$, $BF_{10} = 3.52$, and verbal *n*-back performance, $r = .30$ (95% CI [.038,.520]), $p = .026$, $BF_{10} = 3.70$. In contrast, there was anecdotal evidence against a positive correlation between low verbalisability odours and both visual, $r = .14$ (95% CI [-.124,.391]), $p = .292$, $BF_{10} = 0.49$, and verbal *n*-back performance, $r = .19$ (95% CI [-.080,.429]), $p = .168$, $BF_{10} = 0.77$. Finally, there was strong support for a moderate positive correlation between verbal and visual working memory, $r = .49$ (95% CI [.255,.664]), $p < .001$, $BF_{10} = 371.28$.

- Table 4 about here please -

Discussion

Experiment 4 is the first to compare directly olfactory *n*-back performance for high and low verbalisable odorants against other stimulus types. Consistent with the results of Experiments 1-3, we report superior recognition sensitivity performance for high, compared to low, verbalisable odorants. This effect was replicated despite only using recent-lures, and discarding non-recent-lures, in the calculation of false-alarm rates. That is, the effect is replicated with less statistical power (through the employment of fewer trials) and when using a more valid measure of olfactory working memory (i.e. participants could no longer rely upon relatively weak familiarity signals in order to reject the lures). It is worth noting that despite no longer including non-recent-lures in the false alarm calculation, working memory sensitivity (*A'*) remained significantly above chance for the low verbalisable odorants. This demonstrates, using a more valid measure of working memory, that participants can perform the *n*-back task with low verbalisable odorants. Indeed, it is worth noting that the inclusion of non-recent lures in the *A'* calculation for Experiments 1-3 may have added variance to the data that is unrelated to working memory resources (i.e. reliance on the strength of a familiarity signal).

Absolute *n*-back performance differences between olfactory, visual, and verbal memory (see Figure 8) are difficult to interpret due to the methodological differences across tasks. For example, it is unsurprising that performance on the visual *n*-back task is superior to that of the verbal *n*-back because the former is a 2-back and the latter is a 3-back task. Moreover, presentation times (and the total retention time between *n* presentations of the same item) vary dramatically between the olfactory stimuli and the verbal/visual stimuli. Such methodological differences may affect how the *n*-back tasks are performed and the resources that are employed. Indeed, it is possible that the methodological differences produced task-specific variance that reduced/masked the relationship across *n*–back tasks. However, notwithstanding these limitations, cross-modality correlations were observed that suggest similar memory processes being employed. Consistent with past research (Schmiedek et al. 2014), a strong correlation (*r* = .49) is evident between *n*-back for visual and verbal stimuli. This strong correlation is consistent with the proposed employment of a general (stimulus non-specific) process when performing the *n*-back task (e.g. Schmiedek, Hildebrandt et al., 2009; Schmiedek et al., 2014; Wilhelm, Hildebrandt, & Oberauer, 2013). We show that *n*-back performance for high verbalisable odorants correlate significantly with low verbalisable odorants (*r* = .27), abstract shapes (*r* = .30), and letters (*r* = .30). This relationship is found despite the methodological differences described above. Low verbalisable odorants fail to significantly correlate with either abstract shapes (*r* = .14) or letters (*r* = .19). Our design does not allow us to identify the use of specific strategies in performance of the *n*-back task; however, we argue that shared variance across tasks indicates the employment of common processes irrespective of what those may be. That both verbal and visual *n*-back performance correlates with high verbalisable odorants, suggests common cross-modal processes across stimulus types. This was not found for low verbalisable odorants. We acknowledge that contrasting the absence and presence of these correlations for low and high verbalisable odorants does not of itself mean that the correlations involving high and low verbalisable odorants differ in size (see for example, Nieuwenhuis, Forstmann, & Wagenmakers, 2011). However, the data does indicate substantial evidence for a relationship between high verbalisable odorants and both the visual and verbal stimuli (via the significance test, confidence intervals, and Bayes Factors). There is no evidence in support of these correlations for low verbalisable odorants (with confidence intervals spanning 0 and Bayes Factors anecdotally supporting the null).

Despite low verbalisable odorants not correlating significantly with visual and verbal stimuli, high and low verbalisable odorants *n*-back performance correlated (*r* = .27, although

$BF_{10} < 3$). It is possible that this shared variance reflected methodological similarities (e.g. perceptual experience due to type of stimuli, presentation times etc.), given that these features were unique to the olfactory *n*-back tasks. As noted earlier, direct comparisons between olfactory and visual/verbal *n*-back tasks are confounded by the methodological differences across tasks. However, both high and low verbalisable odorants possess the same methodological disparities with the visual/verbal stimuli, and yet, despite having the same procedure, it is important to note that high verbalisable odorant *n*-back performance correlated with visual and verbal stimuli but low verbalisable odorants did not.

General Discussion

Across a series of four experiments we have shown a consistent *n*-back memory benefit for high, relative to low, verbalisable odorants (replicating the initial finding of Jönsson et al., 2011). A prosaic explanation for this difference was the utilisation of verbal rehearsal for high verbalisable odorants during the inter-trial interval. However, in Experiment 2, opportunities for verbal rehearsal were limited by the inclusion of CA and the performance difference between high and low verbalisable odorants remained. In Experiment 3 we investigated the extent to which *n*-back metacognitive judgments for high and low verbalisable odorants differed. Applying the ‘remember-know’ procedure (Tulving, 1985) to the olfactory *n*-back task, we reported a greater proportion of recollective judgments for high, relative to low, verbalisable odorants. This finding suggests different memory processes being used in the *n*-back task for high and low verbalisable odorants (see also Olsson et al., 2009) (although as noted earlier, these metacognitive differences may be explained in respect to the strengths of the respective memory traces). Consistent with our working assumption that the memory processes for low verbalisable odorants is different to that of other stimulus types, Experiment

4 showed that visual and verbal *n*-back performance correlated with high, but not low, verbalisable odorants.

Experiments 1-4 each produced above chance performance for low verbalisable odorants (see also Jönsson et al., 2011). Whilst we cannot discount the use of impoverished verbal recoding in performance of the task, the use of low verbalisable odorants coupled with a null effect of CA (Experiment 2) suggests that the olfactory *n*-back task does not simply reflect the use of verbal memory. Moreover, Experiment 4 used a more valid measure of working memory by only including recent lures in the analysis and above chance performance remained for the low verbalisable odorants. Indeed, above chance *n*-back performance for low verbalisable odorants is consistent with *n*-back findings for a range of non-verbal stimuli (e.g. Dade et al., 2001; Jaeggi et al., 2010; Schmiedek et al., 2014). Together, these findings demonstrate that the *n*-back task can be successfully performed without reliance upon verbal memory.

That above chance *n*-back performance remains without the employment of verbal rehearsal raises the question of how olfactory representations are maintained within working memory. We suggested that a non-verbal refreshing mechanism could be employed wherein these items are maintained within the episodic buffer (Baddeley, 2012) or indeed an olfactory specific store (Andrade & Donaldson, 2007). Given that refreshing is attentionally demanding and sensitive to secondary tasks (Barrouillet et al., 2004; Barrouillet et al., 2011; Vergauwe et al., 2010), we predicted olfactory *n*-back performance would be impaired by the inclusion of a secondary task. In Experiment 2, we found limited support for this proposition (with only FAs affected by mental rotation). Future work should explore in more depth the effects of cognitive load on the olfactory *n*-back as it may be that the secondary tasks used in the Experiment 2 were insufficiently taxing to disrupt refreshing. Alternatively, it may be that refreshing is not the mechanism used to maintain olfactory representations within working memory.

Notwithstanding the above chance *n*-back performance for low verbalisable odorants, performance for high verbalisable odorants was consistently superior across the four experiments (in line with Jönsson et al., 2011). Our findings in Experiment 2 suggest that this difference is not due to verbal rehearsal during the inter-trial interval, as CA neither affected overall performance or the difference between high and low verbalisable odorants. Analysis of metacognitive judgments in Experiment 3 did, however, suggest differences in the memory processes employed, with more recollection for high verbalisable odorants (see also Larsson et al., 2006; Olsson et al., 2009) and more guessed responses for low verbalisable odorants. Use of recollection has been associated with a high-control working memory strategy (Baddeley, 2012; Barrett et al., 2004; Loaiza et al., 2015) such as maintenance and updating within a ‘rehearsal window’. Whilst recollection responses were not unique to high verbalisable odorants (proportion of correct responses using recollection for high and low verbalisable odorants = .508 and .334, respectively), it is possible therefore that high verbalisable odorants are generally more amenable to working memory processes. However, the findings of Experiment 2 suggest that use of a ‘rehearsal window’ within working memory is not a verbal process, and instead odorants may be maintained within an olfactory store (see Andrade & Donaldson, 2007). In contrast, both a reduction in recollection and an increase in guessed responses for low verbalisable odorants are indicative of a general reduction in memory certainty for these stimuli (Wixted & Mickes, 2010).

It is not clear what drives this difference in memory processes between high and low verbalisable odorants. Whilst we argue that the difference between high and low verbalisable odorants cannot be explained by verbal rehearsal (due to the null effect of CA in Experiment

2), it is possible that generating a meaningful verbal label during presentation of the odorant may facilitate encoding. Indeed, high verbalisable odorants may be characterised by the availability of semantic information, with working memory resources accessing these stored representations (Tomiczek & Stevenson, 2009). Alternatively, the differences between high and low verbalisable odorants may be driven not by verbalisability but a latent covariable. For example, Moss et al. (2016) reported a high correlation between verbalisability and familiarity (r = .88), and it has been argued that highly familiar items require less working memory resources (Oberauer, 2005; Reder et al., 2015).

The correlational analysis of Experiment 4 suggests similarities between high verbalisable odorants and other stimulus types. In contrast, the data suggest that low verbalisable odorants do not correlate with verbal and visual *n*-back performance. These findings are consistent both with previous imagining research showing different patterns of neural activation (Zelano et al., 2009) and qualitative behavioural differences (Moss et al., 2018; Olsson et al., 2009) for high and low verbalisable odorants. Previous research has considered evidence for modularity in olfactory memory (i.e., the employment of a functionally separate olfactory store, e.g. Andrade & Donaldson, 2007; Johnson & Miles, 2009) by examining the extent to which olfactory memory tasks produce qualitatively different patterns of performance relative to other stimulus types. These studies have reported contradictory evidence. That is, some studies show olfactory memory producing patterns of performance consistent with other stimulus types (e.g. Dade et al., 2001; Miles & Hodder, 2005; Miles & Jenkins, 2000; Johnson & Miles, 2007; Moss et al., 2018; White & Treisman, 1997), whereas other studies show olfactory memory to be inconsistent (Johnson & Miles, 2009; Johnson et al., 2013; Moss et al., 2018; Reed, 2000). The current set of results present a possible explanation for such contradictory findings, indicating that high and low verbalisable odorants may be functionally separable (with such division reflected anatomically, Zelano et al., 2009, and behaviourally, Moss et al., 2018). High verbalisable odorants share resources with visual and verbal memory, whereas low verbalisable odorants function differently. Future research should explore further the functional differences and attempt to explain why these differences occur.

In summary, the present set of experiments provides a detailed examination of olfactory *n*-back performance. Building upon the two existing reports of this task (Dade et al., 2001; Jönsson et al., 2011) we show both quantitative and qualitative differences between working memory for high and low verbalisable odorants. Our data support the proposition that memory for low, but not high, verbalisable odorants is qualitatively different to other stimulus types.

References

Andrade, J., & Donaldson, L. (2007). Evidence for an Olfactory Store in Working Memory? *Psychologia*, *50*(2), 76–89.

Annett, J. M., & Leslie, J. C. (1996). Effects of visual and verbal interference tasks on olfactory memory: the role of task complexity. *British Journal of Psychology*, *87*(3), 447–460.

Attneave, F., Arnoult, M. D., & Attneave, F. (1956). The quantitative study of shape and pattern perception. *Psychological Bulletin*, *53*(6), 452–471. https://doi.org/10.1037/h0044049

Baddeley, A. D. (2012). Working Memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*(1), 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Baddeley, A. D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology*, *36A*(2), 233–252. https://doi.org/10.1080/14640748408402157

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 575–589. https://doi.org/10.1016/S0022-5371(75)80045-4

Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual Differences in Working Memory Capacity and Dual-Process Theories of the Mind. *Psychological Bulletin*, *130*(4), 553–573. https://doi.org/10.1037/0033-2909.130.4.553

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General, 133,* 83-100. https://doi: 10.1037/0096-3445.133.1.83

Barrouillet, P., Portrat, S., & Campos, V. (2011). On the law relating processing to storage in working memory. *Psychological Review, 118,* 175-192. https://doi: 10.1037/a0022324

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychologicial Review*, *108*(3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624

Camos, V., & Barrouillet, P. (2014). Attentional and non-attentional systems in the maintenance of verbal information in working memory: the executive and phonological loops. *Frontiers in Human Neuroscience, 8,* https://doi: 10.3389/fnhum.2014.00900

Camos, V., Johnson, M., Loaiza, V., Portrat, S., Souza, A., & Vergauwe, E. (2018). What is attentional refreshing in working memory. *Annals of the New York Academy of Sciences, 1424,* 19-32. https://doi: 10.1111/nyas.13616

Chatham, C. H., Herd, S. A., Brant, A. M., Hazy, T. E., Miyake, A., O'Reilly, R., & Friedman, N. P. (2011). From an executive network to executive control: a computational model of the n-back task. *Journal of Cognitive Neuroscience*, *23*(11), 3598–3619. https://doi.org/10.1162/jocn_a_00047

Chuah, Y. M. L., Maybery, M. T., & Fox, A. M. (2004). The long-term effects of mild head injury on short-term memory for visual form, spatial location, and their conjunction in well-functioning university students. *Brain and Cognition*, *56*(3), 304–312. https://doi.org/10.1016/j.bandc.2004.08.002

Cocchini, G., Logie, R. H., Della Sala, S., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: evidence for domain-specific working memory systems. *Memory & Cognition*, *30*(7), 1086–1095. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12507373

Craig, K. S., Berman, M. G., Jonides, J., & Lustig, C. (2013). Escaping the recent past: Which stimulus dimensions influence proactive interference? *Memory & Cognition*, *41*(5), 650–670. https://doi.org/10.3758/s13421-012-0287-0

Dade, L. A., Zatorre, R. J., Evans, A. C., & Jones-Gotman, M. (2001). Working memory in another dimension: functional imaging of human olfactory working memory. *NeuroImage*, *14*(3), 650–660. https://doi.org/10.1006/nimg.2001.0868

Doty, R. L., Shaman, P., Applebaum, S. L., Giberson, R., Siksorski, L., & Rosenberg, L. (1984). Smell identification ability: changes with age. *Science (New York, N.Y.)*, *226*(4681), 1441–1443. https://doi.org/10.1126/science.6505700

Dudukovic, N. M., Dubrow, S., & Wagner, A. D. (2009). Attention during memory retrieval enhances future remembering. *Memory & Cognition*, *37*(7), 953–961. https://doi.org/10.3758/MC.37.7.953

Dunn, J. C. (2008). The dimensionality of the remember-know task: a state-trace analysis. *Psychological Review*, *115*(2), 426–446. https://doi.org/10.1037/0033-295X.115.2.426

Engen, T., & Ross, B. M. (1973). Long-term memory of odors with and without verbal descriptions. *Journal of Experimental Psychology*, *100*(2), 221–227. https://doi.org/10.1037/h0035492

Evans, L. H., & Wilding, E. L. (2012). Recollection and Familiarity Make Independent Contributions to Memory Judgments. *Journal of Neuroscience*, *32*(21), 7253–7257. https://doi.org/10.1523/JNEUROSCI.6396-11.2012

Frank, R. A., Rybalsky, K., Brearton, M., & Mannea, E. (2011). Odor recognition memory as a function of odor-naming performance. *Chemical Senses*, *36*(1), 29–41. https://doi.org/10.1093/chemse/bjq095

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, *137*(2), 201–225. https://doi.org/10.1037/0096-3445.137.2.201

Ganis, G., & Kievit, R. (2015). *A New Set of Three-Dimensional Shapes for Investigating Mental Rotation Processes: Validation Data and Stimulus Set*. *Journal of Open Psychology Data* (Vol. 3). https://doi.org/10.5334/jopd.ai

Gardiner, J. M., Java, R. I., & Richardson-Klavehn, A. (1996). How level of processing really influences awareness in recognition memory. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *50*(1), 114–122. https://doi.org/10.1037/1196-1961.50.1.114

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (1998). Experiences of remembering, knowing, and guessing. *Consciousness and Cognition*, *7*(7), 1–26. https://doi.org/10.1006/ccog.1997.0321

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (2002). Recognition memory and decision processes: a meta-analysis of remember, know, and guess responses. *Memory (Hove, England)*, *10*(2), 83–98. https://doi.org/10.1080/09658210143000281

Herz, R. S., & Engen, T. (1996). Odor memory: Review and analysis. *Psychonomic Bulletin & Review*, *3*(3), 300–313. https://doi.org/10.3758/BF03210754

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541. https://doi.org/10.1016/0749-596X(91)90025-F

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N -back task as a working memory measure. *Memory*, *18*(4), 394–412. https://doi.org/10.1080/09658211003702171

Jalbert, A., Neath, I., & Surprenant, A. M. (2011). Does length or neighborhood size cause the word length effect? *Memory & Cognition*, *39*(7), 1198–1210. https://doi.org/10.3758/s13421-011-0094-z

Jehl, C., Royet, J., & Holley, A. (1997). Role of verbal encoding in short and long-term odor recognition. *Attention, Perception, & Psychophysics*, *59*(1), 100–110. https://doi.org/10.3758/bf03206852

Johnson, A. J., Cauchi, L., & Miles, C. (2013). Hebbian learning for olfactory sequences. *The Quarterly Journal of Experimental Psychology*, *66*(6), 1082–1089.

https://doi.org/10.1080/17470218.2012.729068

Johnson, A. J., & Miles, C. (2007). Serial position functions for recognition of olfactory stimuli. *Quarterly Journal of Experimental Psychology (2006)*, *60*(10), 1347–1355. https://doi.org/10.1080/17470210701515694

Johnson, A. J., & Miles, C. (2009). Single-probe serial position recall: evidence of modularity for olfactory, visual, and auditory short-term memory. *Quarterly Journal of Experimental Psychology (2006)*, *62*(2), 267–275. https://doi.org/10.1080/17470210802303750

Jönsson, F. U., Møller, P., & Olsson, M. J. (2011). Olfactory working memory: effects of verbalization on the 2-back task. *Memory & Cognition*, *39*(6), 1023–1032. https://doi.org/10.3758/s13421-011-0080-5

Jönsson, F. U., & Olsson, M. J. (2003). Olfactory metacognition. *Chemical Senses*, *28*(7), 651–658. https://doi.org/10.1093/chemse/bjg058

Jönsson, F. U., Tchekhova, A., Lönner, P., & Olsson, M. J. (2005). A metamemory perspective on odor naming and identification. *Chemical Senses*, *30*(4), 353–365. https://doi.org/10.1093/chemse/bji030

Juvina, I., & Taatgen, N. A. (2007). Modeling control strategies in the N-Back task. *8th International Conference on Cognitive Modeling*. Retrieved from http://www.academia.edu/download/30697065/juvina___taatgen.pdf

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. (2007). Working memory, attention control, and the N-back task: a question of construct validity. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *33*(3), 615–622. https://doi.org/2007-06096-010 [pii]\r10.1037/0278-7393.33.3.615

Kärnekull, S. C., Jönsson, F. U., Willander, J., Sikström, S., & Larsson, M. (2015). Long-term memory for odors: Influences of familiarity and identification across 64 days. *Chemical Senses*, *40*(4), 259–267. https://doi.org/10.1093/chemse/bjv003

Katotomichelakis, M., Balatsouras, D., Tripsianis, G., Davris, S., Maroudias, N., Danielides, V., & Simopoulos, C. (2007). The effect of smoking on the olfactory function. *Rhinology*, *45*(4), 273–280.

Koen, J. D., & Yonelinas, A. P. (2016). Recollection, not familiarity, decreases in healthy ageing: Converging evidence from four estimation methods. *Memory*, *24*(1), 75–88. https://doi.org/10.1080/09658211.2014.985590

Larsson, M., Öberg, C., & Bäckman, L. (2006). Recollective experience in odor recognition: Influences of adult age and familiarity. *Psychological Research*, *70*(1), 68–75. https://doi.org/10.1007/s00426-004-0190-9

Loaiza, V. M., & McCabe, D. P. (2012). Temporal–contextual processing in working memory: evidence from delayed cued recall and delayed free recall tests. *Memory & Cognition, 40,* 191–203. https://doi.org/ 10.3758/s13421-011-0148-2.

Loaiza, V. M., Rhodes, M. G., Camos, V., & McCabe, D. P. (2015). Using the process dissociation procedure to estimate recollection and familiarity in working memory: An experimental and individual differences investigation. *Journal of Cognitive Psychology*, *27*(7), 844–854. https://doi.org/10.1080/20445911.2015.1033422

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., & Wagenmakers, E. J. (2015). JASP (Version 0.7) [Computer Software]. Retrieved from https://jasp-stats.org

Lyman, B. J., & McDaniel, M. A. (1986). Effects of encoding strategy on long-term memory for odours. *The Quarterly Journal of Experimental Psychology Section A*, *38*(4), 753–765. https://doi.org/10.1080/14640748608401624

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Miles, C., & Hodder, K. (2005). Serial position effects in recognition memory for odors: a reexamination. *Memory & Cognition*, *33*(7), 1303–1314. https://doi.org/10.1037//0278-7393.26.2.411

Miles, C., & Jenkins, R. (2000). Recency and suffix effects with immediate recall of olfactory stimuli. *Memory*, *8*(3), 195–206. https://doi.org/10.1080/096582100387605

Møller, P., Wulff, C., & Köster, E. P. (2004). Do age differences in odour memory depend on differences in verbal memory? *Neuroreport*, *15*(5), 915–917. https://doi.org/10.1097/01.wnr.0000120561.31269.ef

Monsell, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology*, *10*(4), 465–501. https://doi.org/10.1016/0010-0285(78)90008-7

Morey, R. D., & Rouder, J. N. (2015). BayesFactor (Version 0.9.2) [Computer Software].

Moss, A. G., Miles, C., Elsley, J. V., & Johnson, A. J. (2016). Odorant normative data for use in olfactory memory experiments: Dimension selection and analysis of individual differences. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.01267

Moss, A. G., Miles, C., Elsley, J. V., & Johnson, A. J. (2018). Item-specific proactive interference in olfactory working memory. *Memory, 26*(4), 468-482. doi:10.1080/09658211.2017.1369546

Nguyen, L. A., Ober, B. A., & Shenaut, G. K. (2012). Odor recognition memory: Two encoding trials are better than one. *Chemical Senses*, *37*(8), 745–754. https://doi.org/10.1093/chemse/bjs060

Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E. J. (2011). Erroneous analyses in neuroscience: a problem of significance. *Nature Neuroscience, 14*(9), 1105-1107.

Nijboer, M., Borst, J., van Rijn, H., & Taatgen, N. (2016). Contrasting single and multi-component working-memory systems in dual tasking. *Cognitive Psychology*, *86*, 1–26. https://doi.org/10.1016/j.cogpsych.2016.01.003

Nystrom, L. E., Braver, T. S., Sabb, F. W., Delgado, M. R., Noll, D. C., & Cohen, J. D. (2000). Working Memory for Letters, Shapes, and Locations: fMRI Evidence against Stimulus-Based Regional Organization in Human Prefrontal Cortex. *NeuroImage*, *11*(5 Pt 1), 424–446. https://doi.org/10.1006/nimg.2000.0572

Oberauer, K. (2002). Access to information in working memory: exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 28,* 411–421. https://doi.org/ 10.1037/0278-7393.28.3.411

Oberauer, K. (2005). Binding and inhibition in working memory: individual and age differences in short-term recognition. *Journal of Experimental Psychology. General*, *134*(3), 368–387. https://doi.org/10.1037/0096-3445.134.3.368

Olsson, M. J., Lundgren, E. B., Soares, S. C., & Johansson, M. (2009). Odor memory performance and memory awareness: A comparison to word memory across orienting tasks and retention intervals. *Chemosensory Perception*, *2*(3), 161–171. https://doi.org/10.1007/s12078-009-9051-7

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, *25*(1), 46–59. https://doi.org/10.1002/hbm.20131

Paivio, A. (1990). *Mental representations. A Dual Coding Approach*. Oxford: Oxford University Press.

Pelegrina, S., Lechuga, M. T., Garcia-Madruga, J. A., Elosûa, M. R., Macizo, P., Carreiras, M., … Bajo, M. T. (2015). Normative data on the n-back task for children and young adolescents. *Frontiers in Psychology, 6*:1544. https://doi: 10.3389/fpsyg.2015.01544

Rabin, M. D. (1988). Experience facilitates olfactory quality discrimination. *Perception & Psychophysics*, *44*(6), 532–540. https://doi.org/10.3758/BF03207487

Ragland, J. D., Turetsky, B. I., Gur, R. C., Gunning-Dixon, F., Turner, T., Schroeder, L., … Gur, R. E. (2002). Working memory for complex figures: an fMRI comparison of letter and fractal n-back tasks. *Neuropsychology*, *16*(3), 370–379. https://doi.org/10.1037/0894-4105.16.3.370

Rajaram, S. (1993). Remembering and knowing: two means of access to the personal past. *Memory & Cognition*, *21*(1), 89–102. https://doi.org/10.3758/BF03211168

Ralph, J. (2014). *Statistical manipulation and control strategies of the n-back task.* (Unpublished doctoral thesis). Rensselaer Polytechnic Institute, New York.

Reder, L. M., Liu, X. L., Keinath, A., & Popov, V. (2015). Building knowledge requires bricks, not sand: The critical role of familiar constituents in learning. *Psychonomic Bulletin & Review*, 1–7. https://doi.org/10.3758/s13423-015-0889-1

Redick, T. S., & Lindsey, D. R. B. (2013). Complex span and n-back measures of working memory: a meta-analysis. *Psychonomic Bulletin & Review*, *20*(6), 1102–1113. https://doi.org/10.3758/s13423-013-0453-9

Reed, P. (2000). Serial Position Effects in Recognition Memory for Odors. *Journal of Experimental Psychology: Learning Memory and Cognition*, *26*(2), 411–422. http://doi.org/10.1037/0278-7393.26.2.411

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

Saito, S., Logie, R. H., Morita, A., & Law, A. (2008). Visual and phonological similarity effects in verbal immediate serial recall: A test with kanji materials. *Journal of Memory and Language*, *59*(1), 1–17. https://doi.org/10.1016/j.jml.2008.01.004

Schmiedek, F., Hildebrandt, A., Lövdén, M., Lindenberger, U., & Wilhelm, O. (2009). Complex span versus updating tasks of working memory: the gap is not that deep. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(4), 1089–1096. https://doi.org/10.1037/a0015730

Schmiedek, F., Li, S.-C., & Lindenberger, U. (2009). Interference and facilitation in spatial working memory: age-associated differences in lure effects in the n-back paradigm. *Psychology and Aging*, *24*(1), 203–210. https://doi.org/10.1037/a0014685

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2014). A task is a task is a task: Putting complex span, n-back, and other working memory indicators in psychometric context. *Frontiers in Psychology*, *5*(DEC). https://doi.org/10.3389/fpsyg.2014.01475

Shelton, J. T., Elliott, E. M., Hill, B. D., Calamia, M. R., & Gouvier, W. (2009). A comparison of laboratory and clinical working memory tests and their prediction of fluid intelligence. *Intelligence*, *37*(3), 283–293. https://doi.org/10.1016/j.intell.2008.11.005.A

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science (New York, N.Y.)*, *171*(972), 701–703. https://doi.org/10.1126/science.171.3972.701

Simmons, M. R. (2000). *The central executive and working memory: A dual-task investigation of the n-back task. ProQuest Dissertations and Theses*. ProQuest Information & Learning, US. Retrieved from http://ua.lm.worldcat.org/?genre=article&sid=ProQ:&atitle=&title=The+central+executive+and+working+memory%3A+A+dual-task+investigation+of+the+n-back+task&issn=&date=2000-01-01&volume=&issue=&spage=&author=Simmons%2C+Michelle+R

Smith, E. E., Jonides, J., Koeppe, R. a., Awh, E., Schumacher, E. H., & Minoshima, S. (1995). Spatial versus Object Working Memory: PET Investigations. *Journal of Cognitive Neuroscience*, *7*, 337–356. https://doi.org/10.1162/jocn.1995.7.3.337

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory*, *6*(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174

Stevenson, R. J. (2001). Perceptual learning with odors: implications for psychological accounts of odor quality perception. *Psychonomic Bulletin & Review*, *8*(4), 708–712.

https://doi.org/10.3758/BF03196207

Stevenson, R. J., & Mahmut, M. K. (2013a). The accessibility of semantic knowledge for odours that can and cannot be named. *Quarterly Journal of Experimental Psychology*, *66*(7), 1414–1431. https://doi.org/10.1080/17470218.2012.753097

Stevenson, R. J., & Mahmut, M. K. (2013b). Familiarity influences odor memory stability. *Psychonomic Bulletin & Review*, *20*(4), 754–759. https://doi.org/10.3758/s13423-013-0380-9

Stevenson, R. J., & Wilson, D. A. (2007). Odour perception: An object-recognition approach. *Perception*, *36*(12), 1821–1833. https://doi.org/10.1068/p5563

Tomiczek, C., & Stevenson, R. J. (2009). Olfactory Imagery and Repetition Priming: The Effect of Odor Naming and Imagery Ability. *Experimental Psychology*, *56*(6), 397–408. https://doi.org/10.1027/1618-3169.56.6.397

Tulving, E. (1985). Memory and Consciousness. *Canadian Psychology-Psychologie Canadienne*, *26*(1), 1–12. https://doi.org/10.1037/h0080017

Unsworth, N., & Spillers, G. J. (2010). Variation in working memory capacity and episodic recall: The contributions of strategic encoding and contextual retrieval. *Psychonomic Bulletin & Review*, *17*(2), 200–205. https://doi.org/10.3758/PBR.17.2.200

Vergauwe, E., Barrouillet, P., & Camos, V. (2010). Do mental processes share a domain-general resource? *Psychological Science, 21*(3), 384-390. https://doi: 10.1177/0956797610361340

Vuontela, V., Rämä, P., Raninen, A., Aronen, H. J., & Carlson, S. (1999). Selective interference reveals dissociation between memory for location and colour. *Neuroreport*, *10*(11), 2235–

2240. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10439440

White, T. L., Hornung, D. E., Kurtz, D. B., Treisman, M., & Sheehe, P. (1998). Phonological and perceptual components of short-term memory for odors. *American Journal of Psychology*, *111*(3), 411–434. https://doi.org/10.1111/j.1749-6632.1998.tb10636.x

White, T. L., Møller, P., Köster, E. P., Eichenbaum, H., & Linster, C. (2015). Olfactory Memory. In *Handbook of Olfaction and Gustation* (Vol. 1170, pp. 337–352). https://doi.org/10.1002/9781118971758.ch15

White, T. L., & Treisman, M. (1997). A comparison of the encoding of content and order in olfactory memory and in memory for visually presented verbal materials. *British Journal of Psychology (London, England : 1953)*, *88 (Pt 3)*(February 2016), 459–472. https://doi.org/10.1111/j.2044-8295.1997.tb02651.x

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, *4*(JUL), 1–22. https://doi.org/10.3389/fpsyg.2013.00433

Wilson, D. A., & Stevenson, R. J. (2006). *Learning to smell: olfactory perception from neurobiology to behavior*. Baltimore: JHU Press.

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review*, *117*(4), 1025–1054. https://doi.org/10.1037/a0020874

Yeshurun, Y., Dudai, Y., & Sobel, N. (2008). Working memory across nostrils. *Behavioral Neuroscience*, *122*(5), 1031–1037. https://doi.org/10.1037/a0012806

Yonelinas, A. P. (1999). The contribution of recollection and familiarity contributions to recognition and source memory judgments: A formal dual-process model and an analysis of receiver operating charateristics.pdf. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *25*(6), 1415–1434.

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. https://doi.org/10.1006/jmla.2002.2864

Yonelinas, A. P., & Jacoby, L. L. (1994). Dissociations of processes in recognition memory: Effects of interference and of response speed. *Canadian Journal of Experimental Psychology*, *48*(4), 516–535. https://doi.org/10.1037/1196-1961.48.4.516

Zelano, C., Montag, J., Khan, R., & Sobel, N. (2009). A specialized odor memory buffer in primary olfactory cortex. *PLoS ONE*, *4*(3), e4965–e4965. https://doi.org/10.1371/journal.pone.0004965

Zucco, G. M. (2003). Anomalies in cognition: Olfactory memory. *European Psychologist*, *8*(2), 77–86. https://doi.org/10.1027//1016-9040.8.2.77

Figure Legends

Figure 1. Schematic diagram of the 2-back task. Two buffer items precede the 24 test trials.

Figure 2 (a-c). A' sensitivity (a), proportion of hit rates (b), and proportion of false alarms (c), for low and high verbalisability odorants, across testing blocks. Error bars denote the mean standard error.

Figure 3. Schematic figure of the *n*-back procedure with dual-tasks. Participants were allocated to a group that, during the inter-trial interval, performed one of the following: counting task, mental rotation task, and no concurrent tasks.

Figure 4(a-c). The mean *A'* sensitivity (a), hit rates (b), and false alarms (c), for low and high odorant verbalisability, across the three secondary task groups. Error bars denote the mean standard error.

Figure 5(a-c). Low and high verbalisability odorant *A'* sensitivity (a), hit rates (b), and false alarm rates (c). Error bars denote the mean standard error.

Figure 6(a-b). Proportion of metacognitive response types for hits (a) and false alarms (b) across odorant verbalisability. Error bars denote the mean standard error.

Figure 7. Schematic diagram of verbal 3-back and visual 2-back tasks.

Figure 8. Mean *A'* sensitivity scores across the four different modality *n*-back tasks. Error bars denote the mean standard error.

Table Legends

Table 1.

Odorant verbalisability scores and low or high verbalisability categorisation.

Table 2.

Normative ratings and grouping of olfactory stimuli used in the low and high verbalisability odour n-back tasks. Normative scores for verbalisability (Verb), familiarity (Fam), intensity (Int), pleasantness (Pleas), and hedonic strength (Hed. Str.) are reported.

Table 3

Normative ratings and grouping of olfactory stimuli used in the low and high verbalisability odour n-back tasks. Normative scores for verbalisability (Verb), familiarity (Fam), intensity (Int), pleasantness (Pleas), and hedonic strength (Hed. Str.) are reported.

Table 4

Correlation matrix of A' scores for the four n-back tasks.

Figure 1

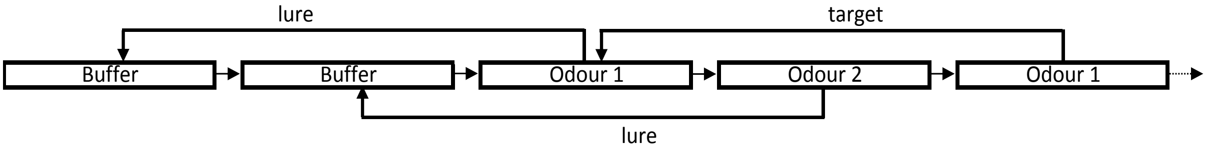

Figure 2(a-c)

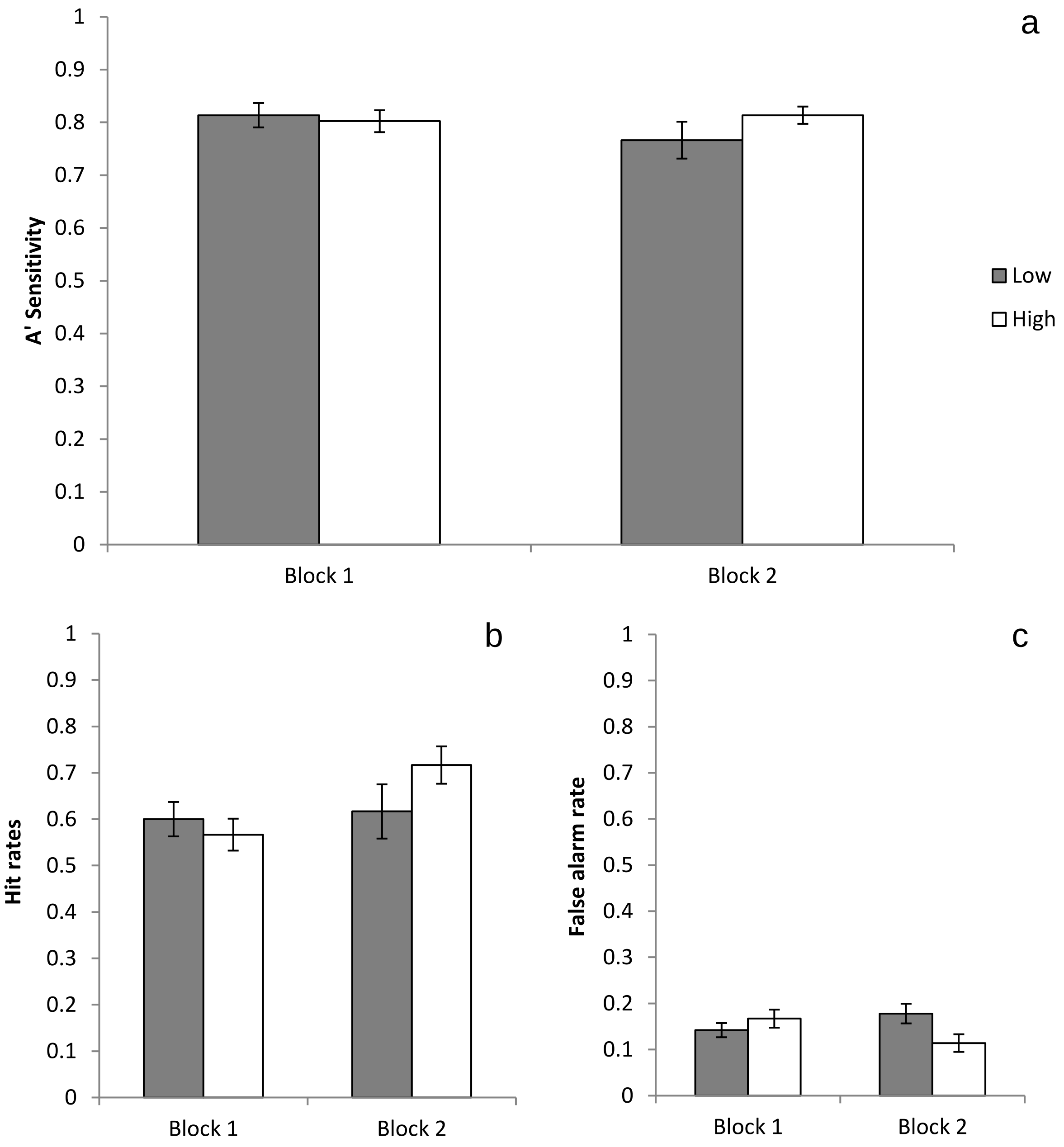
a
A' Sensitivity
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Block 1
Block 2
■ Low
□ High
b
Hit rates
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Block 1
Block 2
c
False alarm rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Block 1
Block 2

Figure 3

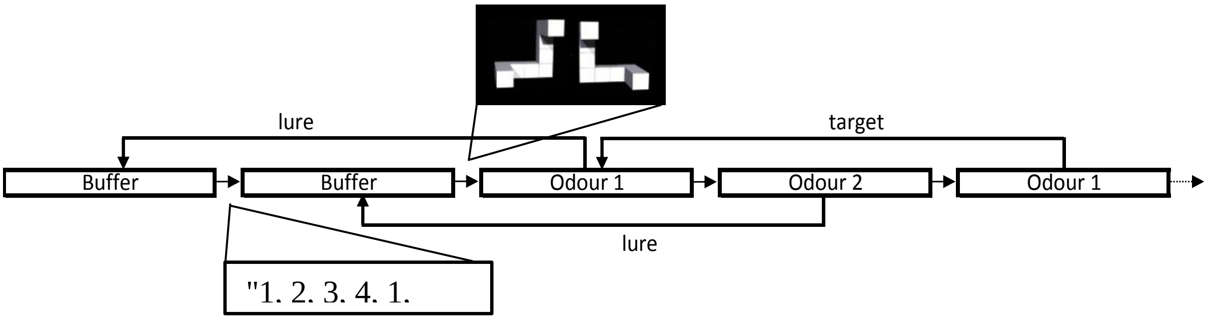

Figure 4(a-c)

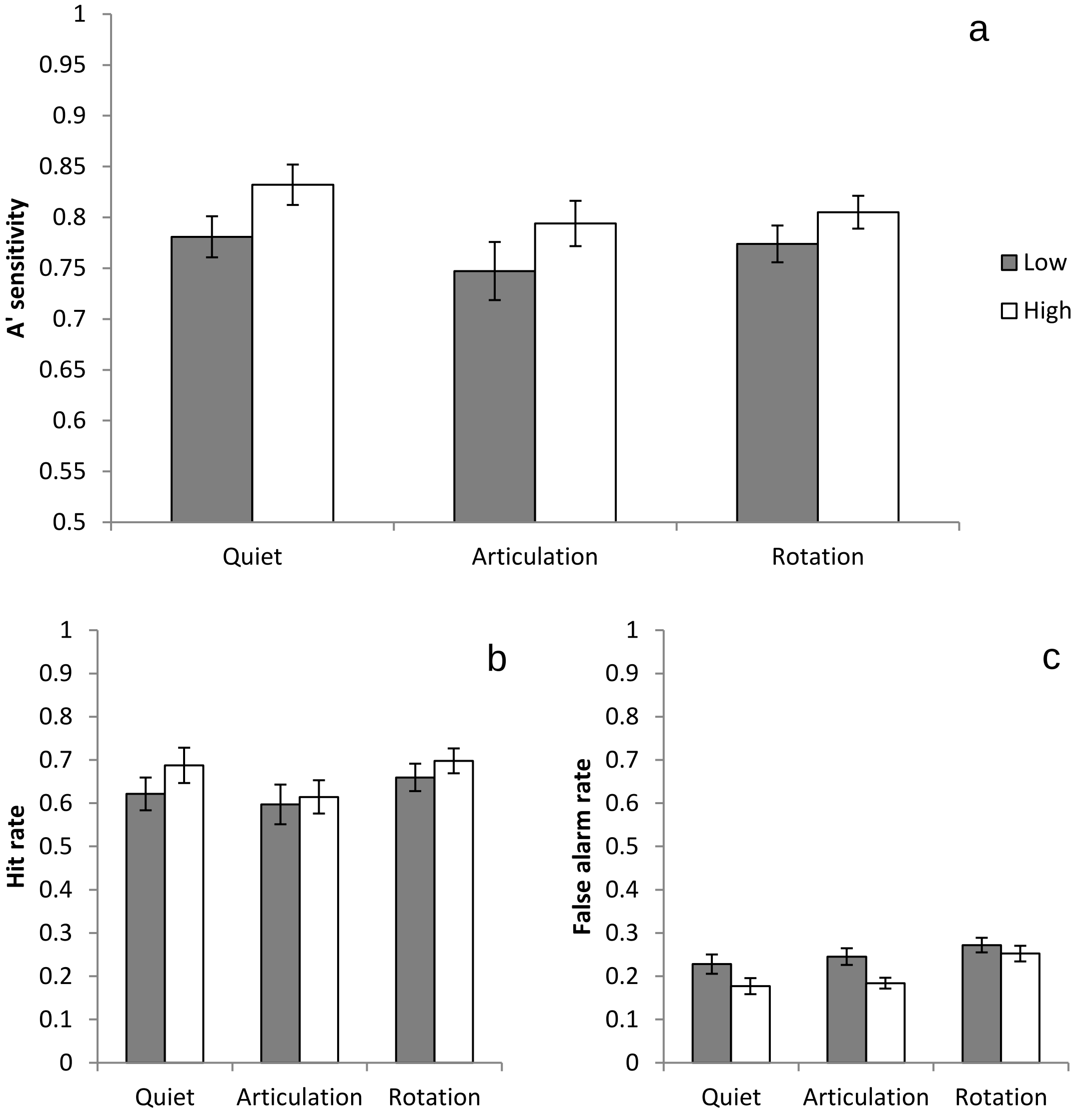
a
A' sensitivity
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
Quiet
Articulation
Rotation
▨ Low
☐ High
b
Hit rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Quiet
Articulation
Rotation
c
False alarm rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Quiet
Articulation
Rotation

Figure 5(a-c)

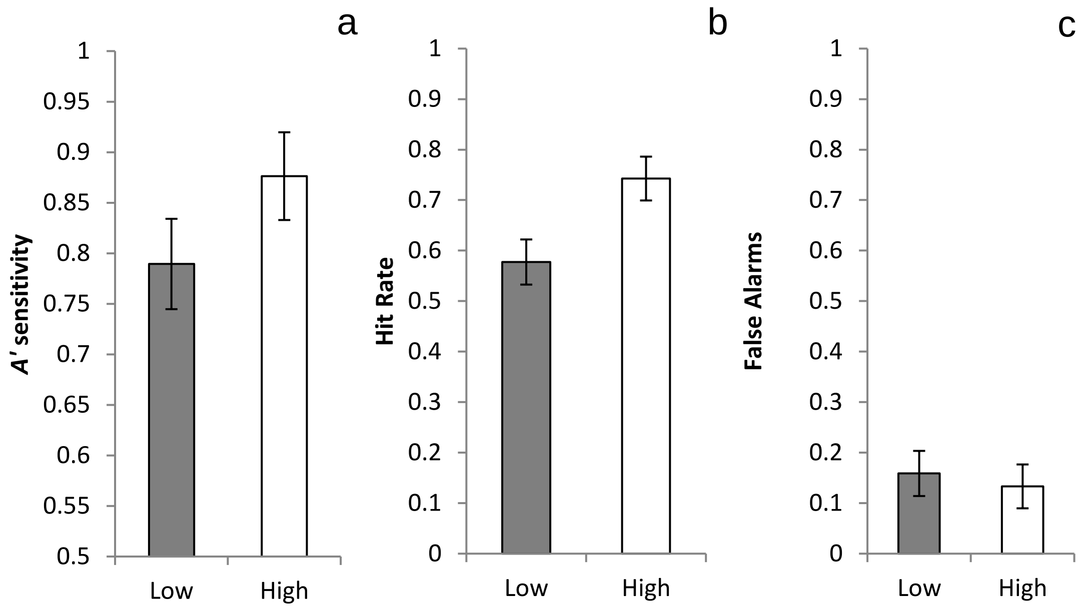
a
b
c
A' sensitivity
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
Low
High
Hit Rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Low
High
False Alarms
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Low
High

Figure 6(a-b)

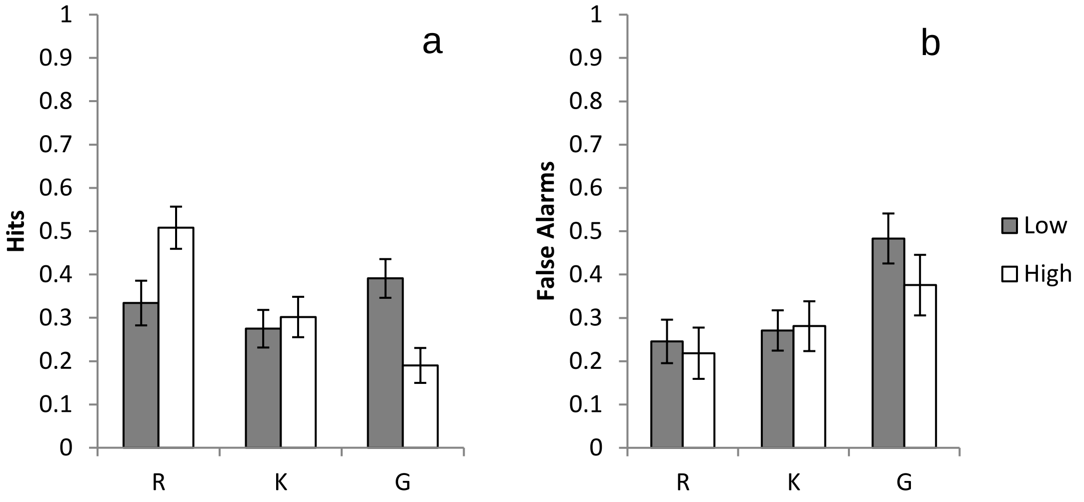

Figure 7

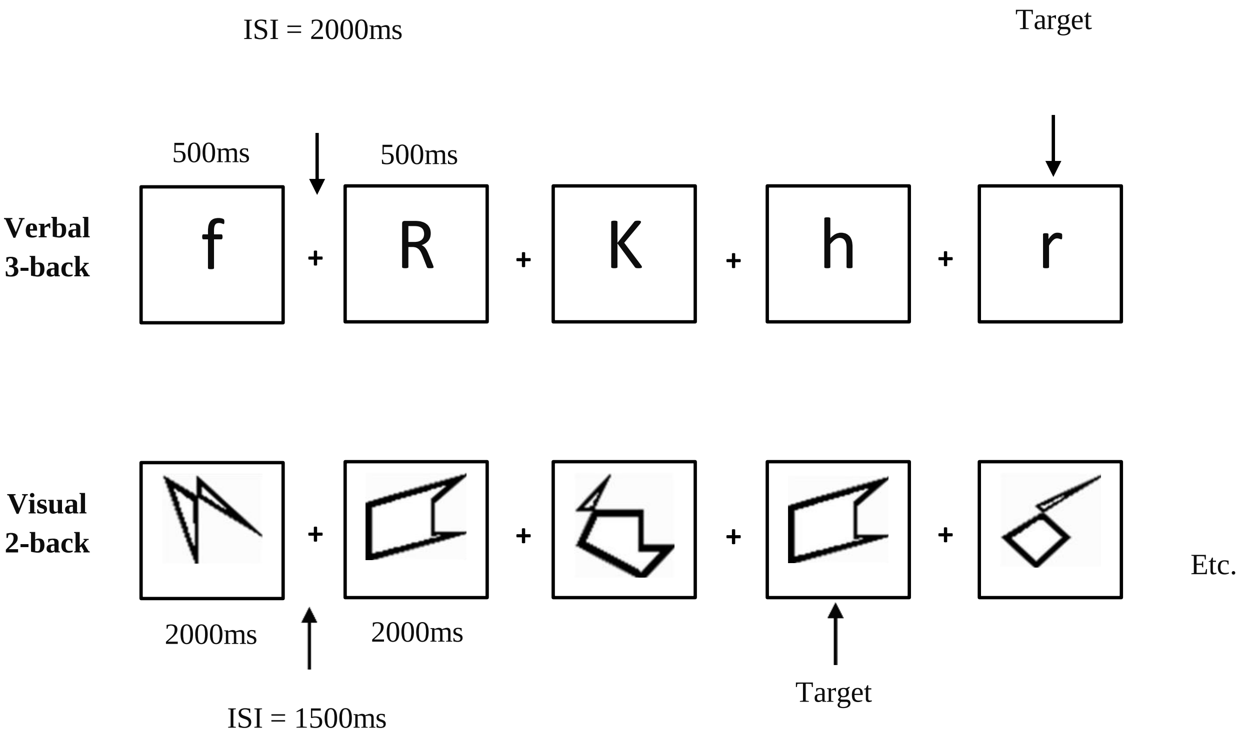
ISI = 2000ms
Target
500ms
500ms
Verbal
3-back
f
+
R
+
K
+
h
+
r
Visual
2-back
+
+
+
+
Etc.
2000ms
2000ms
ISI = 1500ms
Target

Figure 8

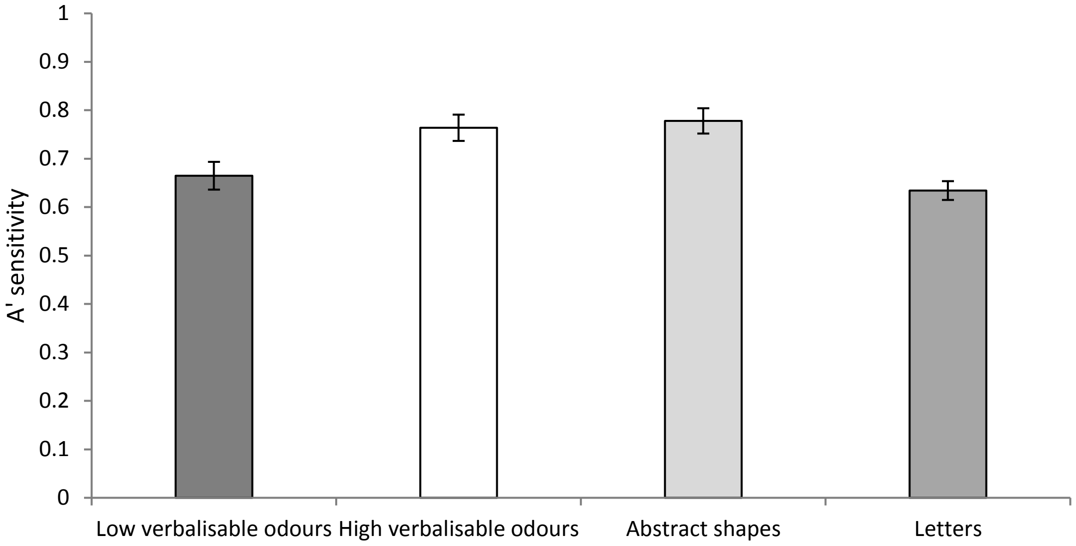
A' sensitivity
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Low verbalisable odours
High verbalisable odours
Abstract shapes
Letters

Table 1

Odour	Verbalisability Category	Verbalisability Score
Carbolic Soap	Low	1.05
Sandalwood	Low	1.17
Nutmeg	Low	1.25
Cuban Cigar Smoke	Low	1.05
Nag Champa*	Low	1.28
Wood Chip	Low	1.18
Cinder Toffee	Low	1.02
Eucalyptus	High	2.45
Garden Mint	High	2.62
Lime	High	2.73
Marzipan	High	2.73
Pear*	High	2.62
Spearmint	High	2.71
Sports Rub	High	2.69

*buffer items not included in n-back analysis.

Table 2

Odour	Task	Verb.	Fam.	Int.	Hed. Str.
Lime	High	2.73	5.70	5.06	1.40
Pear	High	2.62	5.82	5.16	1.40
Eucalyptus	High	2.45	5.88	5.42	1.06
Marzipan	High	2.73	6.12	5.27	1.65
Garden Mint	High	2.62	5.84	5.08	1.66
Sports Rub	High	2.69	5.60	5.52	1.08
Cuban Cigar Smoke	Low	1.05	3.61	5.10	1.33
Sea Shore	Low	1.53	2.96	5.20	1.84
Rum Barrel	Low	1.26	3.10	5.18	1.56
Carbolic Soap	Low	1.05	3.61	5.10	1.33
Patchouli	Low	1.62	3.55	5.06	1.31
Mouse	Low	1.53	3.36	5.06	1.50

Table 3

Odour	Task	Verb.	Fam.	Int.	Hed. Str.
Lime*	High	2.73	5.70	5.06	1.40
Pear	High	2.62	5.82	5.16	1.40
Blackcurrant	High	2.44	5.67	4.85	1.73
Marzipan	High	2.73	6.12	5.27	1.65
Spearmint	High	2.71	5.90	4.96	1.48
Aniseed Balls	High	2.61	5.88	5.40	1.50
Sports Rub	High	2.69	5.60	5.52	1.08
Cheddar Cheese*	Low	1.24	3.14	5.27	1.86
Ginger	Low	1.66	3.39	5.22	1.39
Sea Shore	Low	1.53	2.96	5.20	1.84
Rum Barrel	Low	1.26	3.10	5.18	1.56
Carbolic Soap	Low	1.05	3.61	5.10	1.33
Patchouli	Low	1.62	3.55	5.06	1.31
Mouse	Low	1.53	3.36	5.06	1.50

* Buffer items not included in analysis

Table 4

		1.	2.	3.	4.
1.	Low verbalisable odour 2-back	—	.27*	.14	.19
2.	High verbalisable odour 2-back		—	.30*	.30*
3.	Visual 2-back			—	.49**
4.	Verbal 3-back				—

* *p* < .05

** *p* < .001