

Item-Specific Proactive Interference in Olfactory Working Memory

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

Cognition and Cognitive Neuroscience Research

Centre

Department of Psychology,

Faculty of Science & Technology,

Bournemouth University,

Poole, UK

BH12 5BB

*Correspondence should be sent to: Andrew Moss, Department of Psychology, Bournemouth University, Poole House, Talbot Campus, Fern Barrow, Poole, UK. email: mossa@bournemouth.ac.uk

Word count: 10,684

Item-Specific Proactive Interference in Olfactory Working Memory

We examine item-specific olfactory proactive interference (PI) effects and undertake comparisons with verbal and non-verbal visual stimuli. Using a sequential recent probes task, we show no evidence for PI with hard-to-name odours (Experiment 1). However, verbalisable odours do exhibit PI effects (Experiment 2). These findings occur despite above chance performance and similar serial position functions across both tasks. Experiments 3 and 4 apply words and faces, respectively, to our modified procedure, and show that methodological differences cannot explain the null finding in Experiment 1. The extent to which odours exhibit analogous PI effects to that of other modalities is, we argue, contingent on the characteristics of the odours employed.

Keywords: proactive, interference, olfaction, memory, modality

Memory for a stimulus can be affected by stimuli that both precede (proactive interference: PI) and succeed (retroactive interference: RI) the to-be-remembered (TBR) stimulus. Whilst effects of interference on verbal stimuli are well established (Craig, Berman, Jonides, & Lustig, 2013; Jonides & Nee, 2006; Monsell, 1978; Postman & Underwood, 1973), the role of interference in olfactory memory is both under-researched and contradictory. Early work reporting a flat forgetting function for olfactory stimuli over extended retention intervals (Engen, Kuisma, & Eimas, 1973; see also Jones, Roberts, & Holman, 1978; Engen & Ross, 1973; Lawless & Cain, 1975) was attributed to weak RI coupled with strong PI (Lawless & Engen, 1977). However, this differential weighting in the levels of RI and PI contradict the serial position functions typically reported for olfactory stimuli with odours. That is, strong PI should produce a serial position function with primacy and weak recency. Specifically, monotonically increasing PI throughout the sequence should impair memory for latter list items to a greater extent than early list items. In direct contradiction to that prediction, primacy is rarely observed for olfactory short-term memory tasks (Johnson, Cauchi, & Miles, 2013; Johnson & Miles, 2007, 2009; Miles & Hodder, 2005; c.f. Miles & Jenkins, 2000; Reed, 2000). Indeed, the presence of recency but not primacy (Johnson, Cauchi, & Miles, 2013; Johnson & Miles, 2007; Miles & Hodder, 2005; c.f. Miles & Jenkins, 2000; Reed, 2000), can be interpreted as evidence for RI, with RI monotonically decreasing throughout the sequence (indeed, subsequent support for RI in olfactory memory can be found in Walk & Johns, 1984, and Köster, Degel, & Piper, 2002). Serial position data suggests that olfactory STM is not susceptible to PI and, the present set of experiments, therefore, seek to directly examine this proposition.

There is a paucity of studies examining PI in olfactory memory; with some support for PI found indirectly. For example, Valentin, Dacremont, and Cayeux (2011) showed

recognition memory for odours declined as a function of experimental stage. Whilst this decline in performance may be interpreted as a build-up of PI, it is difficult to deconfound from the more general effects of olfactory fatigue (as reported by Reed, 2000). Köster et al. (2002) examined PI effects more directly using an implicit memory procedure in which two different experimental rooms were paired with odours. At the end of the study, participants were required to rate the extent to which certain odours ‘fit’ 12 different environmental contexts shown on a screen (of which 2 were the rooms used previously). They showed that the paired association for the second room-odour association can be disrupted (as indexed by a reduction in mean rating of ‘fit’) by memory for the initial room-odour association; a demonstration of proactive interference.

The distinction between Valentin et al. (2011) and Köster et al. (2002) is important because it highlights that one can subdivide PI effects into item-nonspecific and item-specific PI (Postle & Brush, 2004; Postle, Brush, & Nick, 2004). Non-specific PI can be conceptualised as a general build-up of interference following repeated exposure to a particular stimulus type and, thus, is difficult to differentiate from olfactory fatigue. In contrast, item-specific interference (e.g. Jonides & Nee, 2006; Monsell, 1978) concerns memory for a previously presented item (e.g. initial presentation of “lavender”) interfering with a subsequent memory for that item (e.g. later presentation of “lavender”). Item-specific PI may, therefore, be taken as a direct measure of PI and is the focus of the present set of experiments.

In the present study, we employ the recent probes task (an established measure of item-specific PI, see Jonides & Nee, 2006, for review). In this task, participants undertake a series of trials, where each trial comprises a TBR memory set that typically numbers 4-items, and is followed by a single yes/no recognition probe. This probe is taken either from the preceding memory set (positive probe), or from an earlier TBR memory set (negative probe).

The important manipulation for this task concerns previous exposure to the negative probe (also referred to as the 'lure'). Negative probes are divided into recent negative (RN) and non-recent negative (NRN). For the RN probes, the probe is taken from the memory set immediately preceding the current trial, whilst for the NRN probes, the probe is taken from the memory set presented 3 trials earlier. Thus, the key manipulation is the recency of the previous presentation of the negative probe. Item-specific PI effects are evidenced by both lengthened reaction times and an increase in errors for the RN probes in comparison to the NRN probes (Monsell, 1978). Both effects are typically interpreted via an increased need to resolve interference (Jonides, Smith, Marshuetz, Koeppe, & Reuter-Lorenz, 1998). That is, due to the strong memory for the RN probe, the individual experiences difficulty in determining whether that strong memory is a consequence of that item being included in the present trial, or being presented in a recent trial. This confusion regarding the origins of the RN probe is also reflected in metacognitive measures. Specifically, confidence ratings for correct rejections of the RN probe are typically lower than those for the NRN probe, and false alarms may reflect the presence of high-confidence intrusion errors (Jacoby, Wahlheim, Rhodes, Daniels, & Rogers, 2010; Wahlheim & Jacoby, 2011).

Different explanations have been proposed to account for these RN probe PI effects. Familiarity-inhibition (Jonides, Badre, Curtis, Thompson-Schill, & Smith, 2002; Mecklinger, Weber, Gunter, & Engle, 2003) states that the RN probe provokes powerful familiarity and this is typically associated with a positive response. Therefore, a correct rejection of the probe necessitates inhibition of that familiarity signal. It is this conflict between familiarity for the RN probe and memory for the TBR items in the present trial that causes errors and lengthened response times (Badre & Wagner, 2005). Similarly, context-retrieval models propose that the RN probe has less accuracy due to errors in the source memory for that item (Badre & Wagner, 2005). Put simply, the recency of the RN probe increases the likelihood

that participants confuse the origins of the probe and believe that it was experienced in the present trial as opposed to the preceding trial.

The present set of experiments examines recent probe PI effects for olfactory stimuli. To date, the extent to which the recent probe PI effect is found across different stimulus types is equivocal (Jonides & Nee, 2006). For example, behavioural recent probe PI effects have been found with abstract symbols, letters, spatial locations, and unfamiliar faces (Badre & Wagner, 2007; Leung & Zhang, 2004; Mecklinger et al., 2003; Postle et al., 2004; Prabhakaran & Thompson-Schill, 2011), but not for colours and some shapes (Postle et al., 2004). It is not clear why these differences occur, but one explanation concerns stimulus distinctiveness (Mecklinger et al., 2003). According to Mecklinger et al. (2003), when stimuli are more distinct the memory representation for the items are more defined (and less fuzzy), resulting in a stronger match/mapping between the RN probe and previous presentation of that item. This stronger match results in an increased PI effect for the RN probe.

A stimulus characteristic pertinent for olfactory stimuli, that may mediate distinctiveness and consequently the recent probe PI effect, is verbalisability. Mecklinger et al. (2003) suggest stimuli that can be represented verbally are more distinct than non-verbal stimuli. This leads to the prediction that stronger PI effects should be observed for those olfactory stimuli that can be easily verbalised. However, previous studies have shown that manipulating levels of verbal facilitation does not affect the recent probe PI effect (Brandon, Jha, Trueswell, Barde, & Thompson-schill, 2003; Brandon 2004, in Jonides & Nee, 2006). Indeed, inhibiting verbal labelling via a concurrent articulation condition has been shown to increase the effects of probe recency (Atkins, Berman, Reuter-Lorenz, Lewis, & Jonides, 2011). The Mecklinger et al. and Atkins et al. studies therefore provide competing predictions for our examination of olfactory PI effects and these competing predictions are directly tested in Experiments 1 and 2. According to the Mecklinger et al. (2003) account, one might predict

stronger PI effects for nameable odours, compared to hard-to-name odours, due to higher levels of distinctiveness when using verbal representations. In contrast, increased interference during articulatory suppression suggests that verbalisation may be a protective factor against recent probe PI effects (Atkins et al., 2011), and therefore one might predict less PI for nameable odours.

Experiment 1 directly tests short-term item-specific PI effects for olfactory memory using the recent probes task. We compare olfactory stimuli that have previously been shown to be difficult to label verbally (Moss, Miles, Elsley, & Johnson, 2016). Hard-to-name odours are initially employed in order to investigate memory for olfactory percepts (rather than verbal labels of those percepts). Indeed, there is evidence that nameable and non-nameable odours may be represented differently in memory (e.g. Zelano, Montag, Khan, & Sobel, 2009), with verbal recoding shown to produce memory effects similar to that shown with words (Olsson, Lundgren, Soares, & Johansson, 2009). To be clear, since verbalisable odours may exhibit PI effects resulting from verbal rather than olfactory perceptions, Experiment 1 employs hard-to-name odours.

In the classical recent probes procedure described above (see Jonides & Nee, 2006), participants are simultaneously presented with an array of TBR items (typically four) and followed by a single test probe. However, since odours cannot easily be differentiated following simultaneous presentation, we describe a modification of the recent probes task in which the TBR items are presented sequentially. In this procedure, participants are presented with sequences of odours followed at test by a single yes/no recognition probe. In the present study we additionally include an analysis of serial position for two reasons. First, this analysis provides another approach through which we can make cross-modal STM comparisons. There is some debate as to whether odours produce qualitatively different serial position functions to that of other stimulus types (e.g. see Johnson & Miles, 2009; Reed, 2000) and

this study assesses cross-modal differences in yes/no recognition functions. Second, immediate yes/no recognition has been shown to produce a specific serial position function for visual stimuli (Hay et al., 2007; Johnson, Volp, & Miles, 2014; Kerr, Avons, & Ward, 1999); in later studies using visual stimuli (Experiments 3 and 4) we are able to check whether our task has produced this canonical function. Based upon these previous studies (Hay et al., 2007; Johnson et al., 2014; Kerr et al., 1999), we predict that recognition for positive probes (i.e. odours that were presented in the preceding sequence) will produce serial position functions comprising recency but not primacy.

It is, however, responses to the negative probes that are important in determining any recent probe PI effects. We therefore compare error rates (false alarms) for the RN probe compared to the NRN probe, with PI evidenced by increased false alarms for the RN probe. In addition, we report response confidence for the test probes as a more subtle measure of PI. Confidence resolution is an item-level gamma correlational measure (Roediger III & DeSoto, 2014) showing the intuitive positive relationship between confidence and accuracy (Wahlheim & Jacoby, 2011). However, one might predict a reduction in the strength of this correlation when responding to a RN probe (Brewer & Sampaio, 2012; Roediger III & DeSoto, 2014; Wahlheim & Jacoby, 2011). Since the strength of familiarity for the RN test probe is of less utility in accurately determining if that item was included in the present TBR memory set, the strength of the correlation between confidence and accuracy should decrease. Consequently, a recent probe PI effect for hard-to-name olfactory stimuli would be evidenced by two outcomes. First, there should be an increased level of false alarms for the RN probe compared to the NRN probe. Second, there should be a reduced accuracy-confidence correlation for the RN probe compared to the NRN probe.

As stated previously, if distinctiveness determines recent probe PI effects (Mecklinger et al., 2003) we predict a lack of PI for the hard-to-name odours. However, if limiting verbal

coding accentuates the recent probe PI effect (see Atkins et al., 2011), we predict a strong PI effect for the hard-to-name odours. Indeed, since parallels have been suggested between the processing of faces and odours (Kärnekull, Jönsson, Willander, Sikström, & Larsson, 2015) we predict a drop in accuracy for the RN probe comparable to that found with faces (Brandon et al., 2003 reported a 15% drop in accuracy for face stimuli).

Experiment 1

Method

Participants

Twenty-four Bournemouth University students (non-smokers, mean age = 25.08, $SD = 5.90$, 11 females, 13 males) participated and received course credit or a £10 honorarium.

Participants who self-reported olfactory impairments (e.g. symptoms of cold) were excluded.

Additionally, participants aged older than 40 years were excluded (due to indications that olfactory-related abilities peak between the third and fifth decade, Doty et al., 1984). The

study received ethical approval via the Bournemouth University research ethics procedure.

Materials

Olfactory stimuli.

The experimental stimuli comprised one-hundred and sixty food and non-food related odorants selected from a corpus of 200 odorants, prepared by Dale Air Ltd.

(www.daleair.com). Each odorant comprised 5ml of an oil-based liquid, and was stored in an opaque test tube in order to mask the odorant's colour. Twenty odours (Appendix 1) were selected as the negative probe items; these odours were purposefully pre-selected as they were previously rated as difficult to verbalise. These ratings were obtained from an earlier

study that collected normative data for 200 odours (Moss et al., 2016). The verbalisability rating used for stimulus selection was scored from 0-3 according to the quality of the verbal labels provided, with a lower score indicating vague or absent verbalisability. The 20 odours selected as negative probe items for the present study scored between 1.20 and 1.84 ($M = 1.61$, $SD = 0.19$), meaning that whilst some verbalisation may be possible, the labels tend to be vague or only refer to a broad category descriptor.

Design

A single yes/no recognition paradigm was employed to investigate the effect of PI on olfactory memory. Participants received 40 trials, where each comprised a sequence of 4 odours followed by a single test-probe. In one half of the trials, the probe was an odour presented in the preceding sequence (positive probe) and in the other half of trials the probe was not presented in the preceding sequence (negative probe). In the 20 positive probe trials (P+), each of the four serial positions were tested on five different trials (i.e. the P+ tested each of the four serial positions an equal number of times). The to-be-remembered sequence was unique, meaning that these items and the corresponding positive probes had not been experienced in any preceding trial. However, the principal independent variable concerned the 20 negative probes (P-): for 10 trials the probe was taken from the trial sequence immediately preceding the present trial (the 'recent negative' probe: RN). For the remaining 10 trials the probe was taken from the trial sequence that occurred three trials prior to the present trial (the 'non-recent negative' probe: NRN). It is these negative probes that were the hard-to-name odours described above.

Three dependent variables were recorded. The yes/no response accuracy was recorded as the number of hits (correct positive probe recognition) and false- alarms (incorrect

negative probe recognition). In addition, ratings were taken of the participant's confidence that their response was correct, ranging from 1 (least confident) to 5 (most confident).

Procedure

Written consent was gained from all participants prior to testing. The memory tasks were performed in a well-ventilated laboratory. Participants sat opposite the experimenter, separated by an obfuscation screen. Throughout testing, participants were instructed to focus on a fixation cross located on this screen to minimise visual interference. Participants received four blocks of ten trials, with each block separated by a 5 minute resting period in which they were able to drink water. For each trial, a sequence of four odours was presented followed by a probe odour. Each odour was presented binocularly for 2-seconds during which participants were instructed to inhale deeply. A 2 second inter-stimulus interval (ISI) separated the presentation of each odour within the sequence, followed by a 5-second retention interval prior to presentation of the test probe. Participants were required to indicate verbally (yes/no) with respect to whether the probe odour had been present in the immediately preceding sequence. Participants additionally provided a verbal confidence rating from 1 (guess) to 5 (certain). Each trial was separated by 3 seconds and the completed task lasted 37 minutes.

Figure 1 provides a schematic of the trial structure and different types of trials. The composition of the sequences and the order of trial types were pseudo-randomised with the following restrictions: (1) In order to maximise the recency of the RN probe, the probe was always taken from the third or fourth serial position of the preceding trial. Consequently, the mean number of intervening items between original presentation of the odour and re-presentation of that odour as the RN probe was 5.5 (in addition, the RN could never follow a trial where that item was also used as the test probe in a positive probe trial). An example is

seen in Figure 1 where the RN probe in trial 3 ('pear') was the fourth item in the TBR sequence of trial 2. In this example, there were 5 odours intervening between original presentation of 'pear' in trial 2 and use of 'pear' as the negative probe in trial 3. (2) The NRN probe was originally presented three trials prior to the current trial, and was taken from the third or fourth serial position of that trial. Consequently, the mean number of intervening items between original presentation of the odour and re-presentation as the NRN probe was 15.5. An example is seen in Figure 1 where the NRN probe in trial 4 ('honey') was the third item in the TBR sequence of trial 1. In this example, there were 16 odours intervening between original presentation of 'honey' in trial 1 and use of 'honey' as the negative probe in trial 4. (3) Presentation of the negative probe and its original presentation in a previous trial could not overlap blocks, since the 5-minute inter-block interval would affect temporal recency of the negative probe. As a result, a block of trials could not begin with a negative probe.

[Figure 1 near here]

Results

Recognition Sensitivity

Signal detection theory was used to determine that correct task performance exceeded chance. Overall response sensitivity (d') was computed using the proportion of hits (H) and false alarms (A), $d' = \Phi^{-1}(H) - \Phi^{-1}(FA)$ (Stanislaw & Todorov, 1999). Perfect scores for hits and correct rejections were adjusted by subtracting $1/(2N)$, where N equals the number of possible hits or correct rejections (Macmillan & Kaplan, 1985). This correction is required because d' is indeterminate for perfect hit or false alarm rates due to an infinite z-score (Stanislaw & Todorov, 1999). Consequently, we assume the 'true' hit or false alarm rate is somewhere between one error and a perfect score, settling on half a miss/false alarm. A one

sample t-test revealed that the d' recognition score ($M = 0.85$, $SD = 0.46$) was significantly above zero (i.e. zero = no sensitivity), $t(23) = 9.07$, $p < .001$, $d = 3.78$, meaning that participants were able to perform the task above chance. In order to assess the possibility of olfactory fatigue (and/or non-specific PI), a one-way analysis of variance (ANOVA) compared d' across the four experimental blocks and revealed a non-significant difference, $F(3, 69) = 0.53$, $p = .664$, $\eta_p^2 = .02$. This indicates a lack of olfactory fatigue and/or non-specific PI.

Serial position analysis

Figure 2a shows the serial position function for hits and reveals a recency advantage. A one-way within-participants ANOVA was conducted and revealed a main effect of serial position, $F(3, 69) = 3.83$, $p = .013$, $\eta_p^2 = .14$. Bonferroni-corrected pairwise comparisons ($\alpha = .017$) revealed a significantly greater number of hits at position 4 compared to position 2. No other comparisons were significant. In contrast to the serial position effects reported for accuracy, confidence ratings (see Figure 2b) did not exhibit a main effect of serial position, $F(3, 63) = 1.13$, $p = .344$, $\eta_p^2 = .05$.

[Figure 2a and 2b near here]

Proactive Interference

Accuracy.

Table 1 displays negative probe accuracy (correct rejections) for NRN and RN probe types. A paired t -test reveals a non-significant difference between NRN ($M = .65$, $SD = .17$) and RN ($M = .64$, $SD = .16$), $t(23) = 0.09$, $p = .464$, one-tailed, $d = 0.02$, 95% CI [-.09, .10]. To further examine whether one can accept the null hypothesis of no PI, we report Bayes Factor (B), calculated using the procedures outlined in Dienes (2014). Proposed cut-offs for

acceptance of a hypothesis (Jeffreys, 1961), states a B above 3 as providing substantial support for the alternative hypothesis, whilst below $1/3$ provides substantial support for the null. A B that falls between $1/3$ and 3 deems the data insensitive as to whether the alternative or null hypothesis should be accepted (where 1 equals equivalent evidence for the null and alternative hypotheses). $B_{H(0, X)}$ refers to the Bayes Factors testing each hypothesis, where ‘H’ indicates a half-normal distribution, ‘X’ the predicted SD of this half-normal, against a null hypothesis of no difference. For the present experiment, the predicted SD was 15% of the NRN probe score, taken from the non-verbal (facial stimuli) recent-probes study in Brandon et al. (2003). The SE calculated from our data was corrected to adjust for a small sample size using the formula $SE * (1 + 20/(df \times df))$ (Dienes, 2008). The B value falls within the ‘insensitive’ range, $B_{H(0, 15\%)} = 0.46$, but does indicate that the null hypothesis (no PI) was 2.19 times more likely than the alternative.

Confidence resolution.

The extent to which confidence judgments are predictive of a correct response for RN and NRN were analysed using confidence resolution. Item-level gamma correlations were calculated separately for RN and NRN probes whereby confidence rating (1-5) was correlated with accuracy (0 or 1) (see Table 1). The coefficients were then compared between the two probe types. A positive gamma coefficient reflects effective ability of confidence ratings to discriminate between correct responses and intrusion errors (Jacoby et al., 2010; Wahlheim & Jacoby, 2011). A reduction in the positive coefficient, or a negative correlation, reflects the influence of intrusion errors where a participant gives a highly confident false alarm. When calculating item-level gamma in the present and future experiments, participant data was not analysed when the responses of a participants were either all correct or all incorrect; this is because it is not possible to calculate a correlation coefficient without variance in scores.

Consistent with the accuracy analysis above, there was no difference between RN and NRN in respect to confidence resolution $t(22) = 0.82$, $p = .211$, one-tailed, 95% CI [-.23, .54], $d = 0.27$. Furthermore, the Bayes Factor (using a predicted effect of 0.24 based on findings in Wahlheim & Jacoby, 2011) was insensitive, $B_{H(0, 0.24)} = 1.11$, indicating a lack of evidence for or against the null and alternative hypotheses.

[Table 1 near here]

Discussion

Experiment 1 used a sequential recent probe task to examine the existence of PI in hard-to-name odours. Recognition sensitivity (d') was significantly above chance, demonstrating that participants were able to perform the task. Yes/no recognition performance for the 4-odour sequences revealed some evidence for recency but no primacy; a function consistent with when visual stimuli are employed in this task (Johnson, Volp, & Miles, 2014; Kerr, Avons, & Ward, 1999). However, the main focus of Experiment 1 concerned analysis of the negative probes (lures) and we report an absence of recent probe PI for hard-to-name odours. This finding is in contrast to the prediction that hard-to-name stimuli exhibit accentuated PI (see Atkins et al., 2011) and historical claims of strong PI in olfactory memory (Lawless & Engen, 1977). This is also in direct contrast to recent probe PI effects found with verbal stimuli and unfamiliar faces (e.g. Brandon et al., 2003; Craig et al., 2013; Postle et al., 2004), but is consistent with the absence of the effect with colours and shapes (Postle et al., 2004).

Importantly, whilst both the accuracy and confidence resolution analyses revealed a non-significant difference between the RN and NRN conditions, Bayes factors showed these comparisons to be insensitive. To be clear, whilst our data clearly fails to support the existence of a difference between RN and NRN (i.e. a recent probe PI effect), our data do not provide unequivocal evidence for the null hypothesis (i.e. that RN and NRN are the same).

Bayes factors revealed that for accuracy the null hypothesis was 2.19 times more likely than the alternative hypothesis, and for confidence resolution, the data supported neither the null or alternative hypothesis ($B = 1.11$). This suggests that there is tentative (although not strong) evidence that there is no difference between RN and NRN in respect to accuracy.

There are a number of explanations as to why PI may be absent in Experiment 1. First, it is possible that the memory task was unsuitable for olfactory stimuli, masking any PI effects. We argue that this is unlikely because (1) performance was significantly above chance and (2) the conventional serial position function was observed for single yes/no recognition (e.g. Johnson et al., 2014; Kerr et al., 1999).

Second, the recent probe PI task typically presents the TBR memory set simultaneously at test (Jonides & Nee, 2006). Due to the constraints of olfactory perception, the memory set was presented sequentially at encoding in Experiment 1. It is possible that this methodological difference precludes PI for two possible reasons. First, in the present study there is a longer interval between initial presentation of the RN lure and its re-presentation as a test probe. Since the RN probe is less temporally recent, it may exhibit less interference. Second, with simultaneous presentation one could conceptualise the TBR items as a single item in memory, whereas for sequential presentation there are 4 discrete items presented. This would result in the RN probe in the present study being less recent in respect to the number of intervening items. The extent to which sequential presentation prevents the recent probe PI effect is addressed in Experiment 3.

A third explanation for the absence of PI in Experiment 1 concerns the characteristics of the stimuli. Our negative probe odours were purposely selected for their low verbalisability score (using normative ratings from Moss et al., 2016). The purpose of this selection was to examine PI effects for olfactory memory rather than memory for the verbal recoding of olfactory stimuli. However, it should be noted that the normative data from Moss et al. (2016)

reveal a strong positive correlation between verbalisability and familiarity ($r = .88$). It therefore follows that the negative probe odours used here possessed low familiarity ratings. Since unfamiliar odours are described as ‘fuzzy’ percepts with overlapping features (Stevenson & Mahmut, 2013a; Wilson & Stevenson, 2006), it is possible that this lack of stimulus distinctiveness may have prevented PI. As described earlier, Mecklinger et al. (2003) argues that for less distinct stimuli, the mapping in memory between the original presentation of the item and its re-presentation as the negative probe is less clear, and, as a result, the level of PI for that item is reduced. This explanation is supported by Prabhakaran and Thompson-Schill (2011) who demonstrated stronger PI effects for RN probes when using famous face stimuli (i.e. more familiar) compared to unfamiliar faces (even after accounting for verbal labelling differences).

Experiment 2

Experiment 2 is designed to examine the extent to which the absence of the PI effect in Experiment 1 is due to using hard-to-name odours. Using nameable odours for the negative probes should increase stimulus distinctiveness and accentuate PI (Mecklinger et al., 2003). Indeed, there is reason to predict differences in memory between nameable and hard-to-name odours, since prior work has shown not only that working memory accuracy levels are higher for nameable odours (e.g. Jönsson, Møller, & Olsson, 2011), but that different areas of the brain are activated for these different odour types (Zelano et al., 2009). In Experiment 2, we expect increased use of verbal codes for our odours; we therefore predict recent probe PI effects that are more in line with that shown for verbal stimuli (e.g. Jonides & Nee, 2006). Specifically, for verbalisable odours we predict item-specific proactive interference effects, such that (1) there are more false positives for the RN probe compared to the NRN probe, and (2) confidence resolution is reduced for the RN probe relative to NRN probe.

Method

Participants

Twenty-four students from Bournemouth University (mean age = 24.96, $SD = 6.37$, females = 15, males = 9) participated in Experiment 2. As in Experiment 1, recruitment criteria required non-smokers under the age of 40 years. None had participated in Experiment 1. The study received ethical approval via the Bournemouth University research ethics procedure.

Materials

The stimuli were as described for Experiment 1, with the crucial difference that 20 odours rated high on verbalisability (Moss et al., 2016) were selected as the negative probe odours (see Appendix 1). The verbalisability score for these odours was 0.78 SD above the mean for the stimulus set. Furthermore, as a manipulation check, an independent t -test revealed that verbalisability score for the negative probe odours used in Experiment 2 was significantly higher than the negative probe odours used in Experiment 1, $t(19) = 13.37, p < .001, d = 4.23$. In addition, the odours used for Experiments 1 and 2 also significantly differed on familiarity, $t(38) = 6.17, p < .001, d = 1.95$.

Additional efforts were made to match the odours used as RN and NRN probes on intensity, pleasantness, and irritability (using the normative values reported by Moss et al., 2016 see Table 2).

[Table 2 near here]

Design

The design followed that described for Experiment 1.

Procedure

The procedure followed that described for Experiment 1.

Results

The same analyses were conducted as described for Experiment 1.

Recognition sensitivity

A one sample t-test revealed that the d' recognition score ($M = 1.05$, $SD = 0.39$) was significantly above zero (i.e. zero = no sensitivity), $t(23) = 13.24$, $p < .001$, $d = 5.52$, meaning that participants were able to perform the task above chance. A one-way analysis of variance (ANOVA) compared d' across the four experimental blocks and revealed a non-significant difference, $F(3, 69) = 0.53$, $p = .672$, $\eta_p^2 = .02$. This, consistent with Experiment 1, indicates a lack of olfactory fatigue and/or non-specific PI.

Serial position analysis

Figure 3a shows the serial position function for hits and, consistent with the findings of Experiment 1, reveals a recency advantage. A one-way within-participants ANOVA was conducted and revealed a main effect of serial position, $F(3, 69) = 7.08$, $p < .001$, $\eta_p^2 = .53$. Bonferroni-corrected pairwise comparisons ($\alpha = .017$) revealed a significantly greater number of hits at position 4 compared to positions 1 and 2. No other comparisons were significant. In contrast to the serial position effects reported for accuracy, confidence ratings (see Figure 3b) did not exhibit a main effect of serial position, $F(3, 63) = 0.67$, $p = .572$, $\eta_p^2 = .03$.

[Figure 3a and 3b near here]

Proactive Interference

Table 3 displays the mean accuracy and confidence resolution for the NRN and RN probes. A paired t -test reveals borderline statistically significant lower accuracy for the RN probes ($M = .60$, $SD = .19$), compared to NRN probes ($M = .70$, $SD = .14$), $t(23) = 2.037$, $p = .027$, one-tailed, $d = 0.85$, 95% CI [.00, .18].

This effect is further supported by the Bayes factor which shows that the alternative hypothesis is 3.96 times greater than the likelihood of the null ($B_{H(0, 15\%)} = 3.96$). These findings indicate a PI effect for recognition accuracy.

Consistent with the accuracy data, confidence resolution was significantly lower for RN probes ($M = -.11$, $SD = .56$), compared to that found with NRN probes ($M = .33$, $SD = .33$), $t(21) = 3.63$, $p = .001$, one-tailed, $d = 0.95$, 95% CI [0.19, 0.69], $B_{H(0, 0.24)} = 103.53$. Taken together, this is strong evidence for a recent probe PI effect with nameable odours.

[Table 3 near here]

Discussion

Experiment 2 reports recent probe PI effects for nameable odours both in respect to the correct rejection of negative probes and in respect to confidence resolution. That is, false alarms were greater for recent (RN) lures compared to NRN lures, and the positive relationship between confidence and accuracy seen for NRN lures was significantly lower, and negative, for RN lures. This is in stark contrast to Experiment 1, where, for hard-to-name odours, we found evidence weighted towards the null hypothesis of no PI effects. In discussing these findings one must firstly consider why hard-to-name and nameable odours differ in PI, and secondly, explore how this speaks to recent probe PI effects with other stimulus types.

First, it is intriguing that nameable and non-nameable odours exhibited different PI effects given similarities in performance across studies. That is, (1) participants performed in both studies above chance, (2) recognition sensitivity did not differ significantly between both tasks ($t(46) = 1.63, p = .110, d = 0.47$), though the data were insensitive to differences ($B_{N(0,0.3)} = 1.18$), and (3) the serial position functions for hits were qualitatively equivalent (recency but no primacy). As outlined earlier, it is possible that stimulus characteristic differences in respect to the distinctiveness of nameable and non-nameable odours, underpinned the difference (Mecklinger et al., 2003). It should be noted that our hard-to-name odours are also less familiar (Moss et al., 2016), and representations for less familiar odours are argued to be less distinct with overlapping features (Stevenson & Mahmut, 2013a; Wilson & Stevenson, 2006). If the negative probe odours are less defined/distinct, it is possible that participants are less aware/confident that the current probe odour maps onto the representation for that same odour in the preceding trial (the RN probe) (an account outlined by Mecklinger et al., 2003). The present findings may therefore not be attributed to stimulus verbalisability specifically, but from a reduction in stimulus distinctiveness from some combination of familiarity and verbalisability differences.

Second, whilst the present findings fit with the Mecklinger et al. explanation, they are not necessarily consistent with recent probe PI effects found with other stimulus effects. For example, Atkins et al. (2011) showed greater PI effects during an articulatory suppression task, though they do suggest this is due to reduced distinctiveness of the episodic information rather than the verbalisability of the stimulus. However, Postle et al. (2004) reported no behavioural evidence for PI with shapes and colours; for which one might expect colours in particular to have easily accessible verbal labels. In addition, PI effects have been found with faces regardless of verbalisation (Brandon et al., 2003); a stimulus that is argued to be processed similarly to that of odours (Kärnekull et al., 2015).

Any attempts at cross-modal comparisons with olfactory PI effects are, however, confounded by methodological differences. To reiterate, our recent probe PI procedure involves sequential presentation of the TBR memory set. In contrast, the classic version of the task involves simultaneous presentation of the memory set (see Craig et al., 2013; Jonides & Nee, 2006). It is therefore possible that differences in the encoding experience of the memory set affects the magnitude of PI. To be clear, whilst Experiment 1 suggests that hard-to-name odours differ from, for example, words (Jonides & Nee, 2006) and faces (Brandon et al., 2003), in not demonstrating a recent probe PI effect, it is possible that this apparent difference is underpinned by differences in method rather than stimulus. The sequential presentation method employed in Experiments 1 and 2 has two important implications. First, the time between initial odour presentation and the recent-negative probe item is greater than those typically seen in the recent-probes task (e.g. Badre & Wagner, 2005; Craig et al., 2013). Second, the relative isolation of the TBR items may have had some unknown effect on the item-specific interference (though it should be noted that Monsell, 1978, did use sequential presentation of verbal stimuli in their seminal work). As a consequence, we need to apply other stimulus types to the sequential recent probe task in order to make meaningful comparisons with the olfactory PI findings reported in Experiments 1 and 2.

Experiment 3

Experiments 3 and 4 apply visual-verbal (words) and visual non-verbal (faces) stimuli to the sequential recent probes tasks used in Experiments 1 and 2. These stimulus types have been shown to exhibit recent probe PI effects when the memory set are presented simultaneously (Brandon et al., 2003; Craig et al., 2013; Postle et al., 2004). If these stimulus types also show PI for the sequential version of the task, it will demonstrate that hard-to-name odours (Experiment 1) differ to other stimulus types in respect to the presence of PI.

Experiment 3 used words¹ and based upon previous work showing PI effects with verbal stimuli (e.g. Jonides & Nee, 2006), we predict PI evidenced by both higher false alarms and lower confidence resolution for the RN probe relative to the NRN probe.

Method

Participants

Twenty-four students from Bournemouth University (mean age = 23.58, *SD* = 9.55, females = 20, males = 4) participated. None had participated in Experiments 1 or 2. The study received ethical approval via the Bournemouth University research ethics procedure.

Materials

Verbal stimuli were 656 high frequency, concrete nouns, selected from the N-Watch (Davis, 2005) default vocabulary of 30,605 words. Mean word length was 6.4 letters (min = 6, max =

¹ Pilot testing using the presentation times employed in Experiments 1 and 2 (2 s exposure time and a 2 s ISI) revealed that direct methodological replication using verbal stimuli is unsuitable due to ceiling effects. Ceiling effects were also produced in two further pilots in which (1) presentation times for each stimulus item were reduced with ISIs increased (thereby maintaining the same temporal interval between re-presentations of the negative probe used in Experiments 1 and 2) and (2) reduced presentation times and ISIs, with increased inter-trial interval increased (again maintaining the same temporal interval between re-presentations of the negative probe used in Experiments 1 and 2). Consequently, presentation times, ISI, and time between trials in Experiment 3 were all proportionately reduced to increase task difficulty (i.e. presentation-interval ratios were identical to Experiments 1 and 2; however absolute timings were reduced). This resulted in a reduction in the time elapsed between initial presentation and re-presentation as a negative probe (a mean of 3.45s between TBR item and RN probe in the present experiment, compared to a mean of 31s for Experiments 1 and 2), meaning the large temporal distance between items as a potential confound was not addressed in this experiment (to preempt, this issue is addressed in Experiment 4). However, the number of items between initial presentation of the item and re-presentation as the negative probe remained identical to that described for Experiments 1 and 2.

7), with 2 syllables. Minimum CELEX frequency (per million words) was 1.62 with a mean of 39.12 ($SD = 88.06$). These words were presented in the centre of a 22 inch monitor in size 18pt. The open-source experimental presentation software OpenSesame (Mathôt, Schreij, & Theeuwes, 2012) was used to present words and record responses.

Design

The design was as described for Experiments 1 and 2.

Procedure

The recent-probes procedure from Experiment 1 and 2 was adapted for verbal stimuli. Due to the brevity of the task, a total of 160 trials were presented to participants, with each trial comprising 4 TBR items followed by a single yes/no recognition probe.

Testing took place at Bournemouth University in an individual laboratory booth. Participants gave written consent, and were instructed on the task procedure. A short (15 trial) practice task preceded the testing phase in order to familiarise participants with the speeded presentation of items. Each TBR item was presented for 100ms, with a 100ms ISI presented as a fixation cross. A 250ms fixation interval separated the final TBR item and presentation of the recognition probe item. When this probe item appeared on the screen, participants were required to make a key press of 'Z' to make a negative response, and a 'V' key press for a positive response. Both decisions were made with the left hand. Immediately following the set membership decision, the probe was removed and 'Confidence?' appeared centrally on the monitor, prompting a numerical confidence rating of 1 (guessing) to 5 (certain) to be made with the right hand on the number keypad. Participants were advised to make their responses to both decisions as quickly and accurately as possible. A 150ms fixation cross separated trials following both responses, and an enforced 30 second break was included every 25 trials. The testing lasted approximately 15 minutes.

Results

To enable direct comparison with Experiments 1 and 2, only the first 20 positive and negative probes (10 RN and 10 NRN) were analysed.

Recognition sensitivity

A one sample t-test revealed that the d' recognition score ($M = 2.35$, $SD = 0.71$) was significantly above zero (i.e. zero = no sensitivity), $t(23) = 16.26$, $p < .001$, $d = 6.78$, meaning that participants were able to perform the task above chance. Item-nonspecific PI was analysed by splitting responses into 10-trial blocks (with the first 5 negative and positive probes allocated to block 1, etc.) and calculating d' . A one-way analysis of variance (ANOVA) compared d' across the four experimental blocks and revealed a non-significant difference, $F(3, 69) = 1.04$, $p = .380$, $\eta_p^2 = .04$. This indicates a lack of task fatigue and/or non-specific PI.

Serial position analysis

Figure 4a shows the serial position function for hits and, consistent with the findings of Experiments 1 and 2, reveals recency but no primacy. A one-way within-participants ANOVA was conducted and revealed a main effect of serial position, $F(3, 69) = 4.64$, $p = .005$, $\eta_p^2 = .17$. Bonferroni-corrected pairwise comparisons ($\alpha = .017$) revealed a significantly greater number of hits at position 4 compared to positions 1 and 2. No other comparisons were significant. As reported for Experiments 1 and 2, serial position effects for confidence ratings (see Figure 4b) were non-significant, $F(3, 63) = 2.01$, $p = .121$, $\eta_p^2 = .08$.

[Figure 4a and 4b near here]

Proactive Interference

Table 4 displays the mean accuracy and confidence resolution for the NRN and RN probes. A paired *t*-test reveals statistically significant lower accuracy for the RN probes ($M = .84$, $SD = .13$), compared to NRN probes ($M = .90$, $SD = .12$), $t(23) = 2.89$, $p = .004$, one-tailed, $d = 0.54$, 95% CI [.02, .11]. Bayes factor was again computed using a predicted drop of 10% for the RN probes; this is based upon verbal recent-probe findings (Craig et al., 2013; Jonides & Nee, 2006; Monsell, 1978). The Bayes factor provided strong support for the alternative hypothesis ($B_{H(0, 7\%)} = 21.06$). These findings indicate a strong PI effect for recognition accuracy.

In contrast to the accuracy data, confidence resolution did not significantly differ between the RN ($M = .14$, $SD = .57$) and NRN probes, ($M = -0.33$, $SD = 0.73$), $t(10) = 1.76$, $p = .055$, one-tailed, $d = 0.72$, 95% CI [-0.13, 1.07], $B_{H(0, 0.7\%)} = 1.91$.

[Table 4 near here]

Discussion

Experiment 3 provides strong support for recent probe PI effects for words using the sequential presentation method employed in Experiments 1 and 2. This strong effect was found in respect to accuracy. It is, however, also worth noting that PI effects were not found for words in respect to confidence resolution. This could be explained by low statistical power for this analysis. That is, calculating confidence resolution necessitates that the participant provides both correct and incorrect responses for the RN and NRN probes. However, since 13 participants provided perfect performance for the RN and/or the NRN probes, this prevented confidence resolution from being calculated for those participants. As a result, the statistics are based upon a sample of 9. Nevertheless, it should be noted that for the analysed data, the effect size was large ($d = 0.72$), providing some support for a PI effect

with confidence resolution.

The findings of Experiment 3, therefore, demonstrate that the apparent difference in susceptibility to PI for odours (Experiment 1) and verbal stimuli (Jonides & Nee, 2006) cannot be explained by the sequential presentation of the TBR items. Here we have shown that the effect remains for verbal stimuli when the TBR items are presented sequentially rather than simultaneously. It is, however, of interest to note that whilst PI was absent for Experiment 1 but observed for words in the present study, both hard-to-name odours and words exhibited qualitatively equivalent serial position functions (recency but no primacy: a function consistent with previous single yes/no recognition studies with verbal stimuli, McElree & Doshier, 1989; Monsell, 1978). This shows some similarity in the memory functioning of hard-to-name odours and words consistent with other recognition tasks (e.g. Johnson & Miles, 2007).

Notwithstanding the use of sequential presentation in Experiments 1 and 3, there remains an important methodological difference between the two experiments. To avoid ceiling effects, the words were presented at a faster rate (100ms with a 100ms ISI) than the odours (2s with a 2s ISI). As a result, the average interval between initial presentation of an item and its re-presentation as the NRN or RN probe differed dramatically for words (NRN interval = 9.65s, RN interval = 3.45s) compared to odours (NRN interval = 79s, RN interval = 31s). If one assumes that PI reduces over time (an assumption on which the recent probe task is premised), then Experiment 1 is weighted against observing a PI effect for hard-to-name odours relative to words in Experiment 3. Consequently, the difference between hard-to-name odours and words in respect to PI may be due to presentation intervals rather than stimuli per se.

Experiment 4

In Experiment 4 we address the criticism that PI effects were not found for hard-to-name odours (Experiment 1) due to long presentation intervals between initial presentation of the item and its re-presentation as the negative probe. For this experiment we employ faces as TBR stimuli for three reasons: (1) faces have previously exhibited a recent probe PI effect (Brandon et al., 2003), (2) pilot work revealed performance not to be at the ceiling when timings more closely reflected that used in Experiments 1 and 2, and (3) faces are often considered a suitable comparison stimuli for odours due to possible holistic processing for both classes of stimuli (Kärnekull et al., 2015; Stevenson & Mahmut, 2013b). To ensure that the presentation intervals matched that described in Experiments 1 and 2, a 24 s interval was introduced between each trial. As in the previous experiments, PI is examined by comparison between the RN and NRN probes in respect to both accuracy and confidence resolution.

Method

Participants

Twenty-four students from Bournemouth University (mean age = 20.76, $SD = 4.88$, females = 22, males = 2) participated. None had participated in Experiments 1-3. The study received ethical approval via the Bournemouth University research ethics procedure.

Materials

One-hundred and sixty faces were randomly selected from the Glasgow Unfamiliar Face Database (Burton, White, & McNeill, 2010). All .jpg images were 350px x 473px, showed a front (full face) view with a neutral expression, and were cropped to remove any visible clothing. Stimuli were presented in the centre of a 22-inch 60Hz monitor and responses collected using OpenSesame (Mathôt et al., 2012).

Design

The same design was used as described for Experiments 1-3.

Procedure

Participants were tested individually in a laboratory booth at Bournemouth University.

Participants were presented with on-screen instructions, and initiated the task by pressing the space bar. The trial started 1s after initiating the task, where four TBR items were displayed sequentially for 250ms, with an ISI of 250ms. A retention interval of 625ms separated the final TBR item and presentation of the recognition probe. The response format followed that described for Experiment 3. Following these responses, a 24s delay (required to approximately match the total time of 29 seconds between initial presentation in the fourth serial position and RN probe, in Experiments 1 and 2) separated trials. Participants were warned 1 second before the beginning of the next trial by a change in the number of fixation dots in the centre of the screen. A total of 40 trials were performed, matching the 20 positive and 20 negative probes presented in Experiments 1 and 2. A 1-minute break separated each block of 10 trials.

Results

Recognition sensitivity

A one sample t-test revealed that the d' recognition score ($M = 2.22$, $SD = 0.45$) was significantly above zero (i.e. zero = no sensitivity), $t(23) = 24.19$, $p < .001$, $d = 10.09$. A one-way analysis of variance (ANOVA) compared d' across the four experimental blocks and revealed a non-significant difference, $F(3, 69) = 0.91$, $p = .441$, $\eta_p^2 = .04$. This indicates a lack of task fatigue and/or item non-specific PI.

Serial position analysis

Figure 5a shows the serial position function for hits and, consistent with the findings of Experiments 1-3, reveals recency but no primacy. A one-way within-participants ANOVA was conducted and revealed a main effect of serial position, $F(3, 69) = 3.37, p = .023, \eta_p^2 = .13$. Bonferroni-corrected pairwise comparisons ($\alpha = .017$) revealed a significantly greater number of hits at position 4 compared to position 2. No other comparisons were significant. Confidence ratings (see Figure 5b) also revealed a significant main effect of serial position, $F(3, 63) = 13.17, p < .001, \eta_p^2 = .36$, with a strong recency effect evidence (position 4 > positions 1-3).

[Figure 5a and 5b near here]

Proactive Interference

Table 5 displays the mean accuracy and confidence resolution for the NRN and RN probes. A paired t -test reveals statistically significant lower accuracy for the RN probes ($M = .70, SD = .16$), compared to NRN probes ($M = .80, SD = .16$), $t(23) = 3.09, p = .003$, one-tailed, $d = 0.63$, 95% CI [.03, .16]. Bayes factor was again computed using a predicted drop of 15% for the RN probes (based on the recent probe PI effect reported for faces, Brandon et al., 2003). The Bayes factor provided strong support for the alternative hypothesis ($B_{H(0, 15\%)} = 32.43$). These findings indicate a strong PI effect for recognition accuracy.

In contrast to the accuracy data, confidence resolution did not significantly differ between the RN ($M = .03, SD = .42$) and NRN probes, ($M = 0.13, SD = 0.56$), $t(18) = 0.67, p = .256$, one-tailed, $d = 0.22$, 95% CI [-0.23, 0.44], $B_{H(0, 0.24)} = 0.92$.

[Table 5 near here]

Discussion

Experiment 4 provides a demonstration of recent probe PI with non-verbal visual stimuli (faces) using the sequential presentation method of the TBR items. This demonstrates that the recent probe PI effect found with faces (Brandon et al., 2003) is not reliant on simultaneous presentation at encoding. In addition, the PI effect was found despite lengthened intervals between presentation of an item and its re-presentation as a negative probe. A 24s interval was introduced between the trials to ensure parity with Experiment 1 in respect to the timings of each trial. Following this manipulation we are able to conclude that, in contrast to faces, hard-to-name odours exhibit no recent probe PI effect, and that this difference is neither a result of sequential presentation of the TBR items nor due to lengthened intervals between the re-presentation of items.

It should, however, be noted that PI was not reported for confidence resolution. Unlike Experiment 3, this non-significant effect does not appear to be due to reduced statistical power. Indeed, the effect size ($d = 0.22$) for the confidence interval resolution was notably smaller than Experiments 2 and 3 ($d = 0.95$ and 0.72 , respectively). It is unclear why in Experiment 4, accuracy but not confidence resolution supports PI.

General Discussion

Across four experiments we have examined cross-modal PI effects using the recent probes task. We have shown evidence for PI with nameable odours (Experiment 2), words (Experiment 3), and faces (Experiment 4), but no PI for hard-to-name odours (Experiment 1). Such a finding may provide support against a domain-general effect in the recent-probes task (Jonides & Nee, 2006; Leung & Zhang, 2004). In addition, our findings with words and faces extend the established recent probes effect (Brandon et al., 2003; Craig et al., 2013; Jonides & Nee, 2006) to a version of the task in which the TBR memory set is presented sequentially

(Experiments 3 and 4).

It is unclear why hard-to-name odours have not demonstrated recent probe PI effects. However, the contrasting PI findings of Experiments 1 (hard-to-name odours) and 2 (nameable odours) are consistent with other studies showing memory differences for nameable and hard-to-name odours (e.g. Jönsson et al., 2011; Moss, Miles, Elsley, & Johnson, in prep; Zelano et al., 2009). A prosaic explanation for the current data is that olfactory stimuli do not elicit item-specific PI effects, and that the PI effects in Experiment 2 are illustrative of verbal memory following (at least partial) verbal recoding of the odours. Alternatively, the different effects may be due to stimulus characteristics that correlate with stimulus verbalisability. The extent to which odours in the current are nameable correlates strongly with familiarity ($r = .88$; Moss et al., 2016). As noted earlier, it is possible that the reduced distinctiveness of these unfamiliar odours resulted in disrupted matching in memory between the original presentation of the item and its re-presentation as the negative probe (Mecklinger et al., 2003). This reduced matching would attenuate PI.

It should be noted that the present findings contradict that of Köster et al. (2002) who examined PI for odours using implicit memory. In their study, PI for the incidental associations between odours and rooms was only found when the odours were not identified. In contrast, the present study only reported PI for nameable odours. It is difficult to make such cross-study comparisons, but these findings may suggest that the verbalisability of the odour has different effects for explicit and implicit memory tasks. Moreover, it is suggested by Köster et al. that such explicit memory tasks may be unsuitable for odours since the experience and learning of odours in everyday life is epiphenomenal (i.e. the present odour is incidentally associated with the present experience). Consequently, when odours are tested explicitly it encourages participants to employ verbal coding, with the task becoming a de facto measure of verbal memory. However, we argue that this issue with explicit odour

memory tasks has been mitigated through the employment of hard-to-name odours in Experiment 1 (odours previously identified as hard-to-name in a large scale normative study, Moss et al., 2016). Indeed, the observation that the hard-to-name odours in Experiment 1 produced the canonical serial position curve supports the applicability of odours to explicit memory tasks typically used with verbal and visual stimuli. Moreover, whilst it remains possible that participants attempt to employ rudimentary labels for these odours, we suggest that the differences between Experiment 1 and 2 in respect to the presence of PI suggest that: (1) hard-to-name and nameable odours are affected differently by PI, and (2) nameable odours may well be represented as verbal code.

If one accepts that (non-verbal) olfactory stimuli are not susceptible to item-specific PI, it suggests qualitative differences between olfactory memory and other stimulus types. This difference is curious given similarities in the serial position functions produced across stimuli in the present study for single yes/no recognition (recency but not primacy, consistent with Hay, Smyth, Hitch, & Horton, 2007; Johnson et al., 2014; Kerr et al., 1999). However, beyond the present study, memory for olfactory stimuli have produced serial position functions that appear qualitatively different to other stimulus types (Johnson et al., 2013; Johnson & Miles, 2009; Reed, 2000). These qualitative differences for olfactory memory add credence to the proposal that olfactory stimuli are represented within a modality specific store (Andrade & Donaldson, 2007), but that verbal dual-coding of the stimuli can supplement the perceptual representation (Paivio, 1990; Yeshurun, Dudai, & Sobel, 2008).

The present study demonstrates that the characteristics of the olfactory stimuli can determine the reported memory effects (a claim also made for other stimulus types, e.g. Hay et al., 2007; Horton, Hay, & Smyth, 2008). This finding adds weight to the proposal by Moss et al. (2016) that any attempts at investigating whether olfactory memory operates

analogously to other stimulus types must consider the characteristics of the odours used in that study.

Conclusion

In summary, the present study demonstrates item-specific PI effects in verbal memory, visual (face) memory, and in memory for familiar (verbalisable) olfactory stimuli. However, these effects are not found with hard-to-name odours. Our study therefore adds weight to the proposition that differences exist in terms of how nameable and non-nameable odours are represented (Zelano et al., 2009) and suggest that making comparisons between olfactory memory and that of other stimulus types might be affected by the choice of odours; a control that requires further consideration in future work (see Moss et al., 2016).

Acknowledgements

We thank Jamie Goodliffe for his assistance with programming Experiments 3 and 4.

References

- Andrade, J., & Donaldson, L. (2007). Evidence for an Olfactory Store in Working Memory? *Psychologia*, *50*(2), 76–89.
- Atkins, A. S., Berman, M. G., Reuter-Lorenz, P. a, Lewis, R. L., & Jonides, J. (2011). Resolving semantic and proactive interference in memory over the short-term. *Memory & Cognition*, *39*(5), 806–817. <https://doi.org/10.3758/s13421-011-0072-5>
- Badre, D., & Wagner, A. D. (2005). Frontal lobe mechanisms that resolve proactive interference. *Cerebral Cortex*, *15*(12), 2003–2012. <https://doi.org/10.1093/cercor/bhi075>
- Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, *45*(13), 2883–2901. <https://doi.org/10.1016/j.neuropsychologia.2007.06.015>
- Brandon, M., Jha, A. P., Trueswell, J. C., Barde, L. H. F., & Thompson-schill, S. L. (2003). Proactive Interference in Verbal and Non-Verbal Working Memory. In *the Annual Meeting of the Psychonomic Society*. Vancouver, Canada.

- Brewer, W. F., & Sampaio, C. (2012). The metamemory approach to confidence: A test using semantic memory. *Journal of Memory and Language*, *67*(1), 59–77.
<https://doi.org/10.1016/j.jml.2012.04.002>
- Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*, *42*(1), 286–291. <https://doi.org/10.3758/BRM.42.1.286>
- 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>
- Davis, C. J. (2005). N-watch: a program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, *37*(1), 65–70.
<https://doi.org/10.3758/BF03206399>
- Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. Palgrave Macmillan. Palgrave Macmillan.
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*(July), 1–17. <https://doi.org/10.3389/fpsyg.2014.00781>
- 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>
- Engen, T., Kuisma, J. E., & Eimas, P. D. (1973). Short-term memory of odors. *Journal of Experimental Psychology*, *99*(2), 222–225. <https://doi.org/10.1037/h0034645>
- Hay, D. C., Smyth, M. M., Hitch, G. J., & Horton, N. J. (2007). Serial position effects in short-term visual memory: A SIMPLE explanation? *Memory & Cognition*, *35*(1), 176–190. <https://doi.org/10.3758/BF03195953>
- Horton, N., Hay, D. C., & Smyth, M. M. (2008). Hebb repetition effects in visual memory: the roles of verbal rehearsal and distinctiveness. *Quarterly Journal of Experimental Psychology (2006)*, *61*(12), 1769–77. <https://doi.org/10.1080/17470210802168674>
- Jacoby, L. L., Wahlheim, C. N., Rhodes, M. G., Daniels, J. A., & Rogers, C. S. (2010). Learning to diminish the effects of proactive interference: Reducing false memory for young and older adults. *Memory & Cognition*, *38*(6), 820–829.
<https://doi.org/10.1080/17439760500510569>.Nice

- Jeffreys, H. (1961). *The Theory of Probability*. OUP Oxford.
- 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), 1347–55.
<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>
- Johnson, A. J., Volp, A., & Miles, C. (2014). Immediate recognition memory for wine. *Journal of Cognitive Psychology*, *26*(2), 127–134.
<https://doi.org/10.1080/20445911.2013.869225>
- Jonides, J., Badre, D., Curtis, C., Thompson-Schill, S. L., & Smith, E. E. (2002). Mechanisms of Conflict Resolution in Prefrontal Cortex. *Principles of Frontal Lobe Function*.
<https://doi.org/10.1093/acprof:oso/9780195134971.003.0015>
- Jonides, J., & Nee, D. E. (2006). Brain Mechanisms of Proactive Interference in Working Memory. *Neuroscience*, *139*(1), 181–193.
<https://doi.org/10.1016/j.neuroscience.2005.06.042>
- Jonides, J., Smith, E. E., Marshuetz, C., Koeppe, R. A., & Reuter-Lorenz, P. A. (1998). Inhibition in verbal working memory revealed by brain activation. *Proc Natl Acad Sci U S A*, *95*(14), 8410–8413. <https://doi.org/10.1073/pnas.95.14.8410>
- 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>
- 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>
- Kerr, J. R., Avons, S. E., & Ward, G. (1999). The effect of retention interval on serial position curves for item recognition of visual patterns and faces. *Journal of*

- Experimental Psychology: Learning, Memory, and Cognition*, 25(6), 1475–1494.
<https://doi.org/10.1037//0278-7393.25.6.1475>
- Köster, E. P., Degel, J., & Piper, D. (2002). Proactive and retroactive interference in implicit odor memory. *Chemical Senses*, 27(3), 191–206.
<https://doi.org/10.1093/chemse/27.3.191>
- Lawless, H. T., & Engen, T. (1977). Associations to odors: interference, mnemonics, and verbal labeling. *Journal of Experimental Psychology: Human Learning and Memory*, 3(1), 52–59. <https://doi.org/10.1037/0278-7393.3.1.52>
- Leung, H.-C., & Zhang, J. X. (2004). MUY BUENO Interference resolution in spatial working memory. *NeuroImage*, 23(3), 1013–9.
<https://doi.org/10.1016/j.neuroimage.2004.07.053>
- Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, 98(1), 185–99. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/4034817>
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44(2), 314–24.
<https://doi.org/10.3758/s13428-011-0168-7>
- McElree, B., & Doshier, B. A. (1989). Serial Position and Set Size in Short-Term Memory : The Time Course of Recognition. *Journal of Experimental Psychology: General*, 118(4), 346–373.
- Mecklinger, A., Weber, K., Gunter, T. C., & Engle, R. W. (2003). Dissociable brain mechanisms for inhibitory control: effects of interference content and working memory capacity. *Cognitive Brain Research*, 18(1), 26–38.
<https://doi.org/http://dx.doi.org/10.1016/j.cogbrainres.2003.08.008>
- Miles, C., & Hodder, K. (2005). Serial position effects in recognition memory for odors: a reexamination. *Memory & Cognition*, 33(7), 1303–14. <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>
- 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](https://doi.org/10.1016/0010-0285(78)90008-7)

- 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>
- 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>
- Paivio, A. (1990). *Mental representations. A Dual Coding Approach*. Oxford University Press.
- Postle, B. R., & Brush, L. N. (2004). The neural bases of the effects of item-nonspecific proactive interference in working memory. *Cogn Affect Behav Neurosci, 4*(3), 379–392. <https://doi.org/10.3758/CABN.4.3.379>
- Postle, B. R., Brush, L. N., & Nick, A. M. (2004). Prefrontal cortex and the mediation of proactive interference in working memory. *Cognitive, Affective & Behavioral Neuroscience, 4*(4), 600–608. <https://doi.org/10.3758/CABN.4.4.600>
- Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory & Cognition, 1*(1), 19–40. <https://doi.org/10.3758/BF03198064>
- Prabhakaran, R., & Thompson-Schill, S. L. (2011). The Price of Fame: The Impact of Stimulus Familiarity on Proactive Interference Resolution. *Journal of Cognitive Neuroscience, 23*(4), 816–831. <https://doi.org/10.1162/jocn.2010.21501>
- Reed, P. (2000). Serial Position Effects in Recognition Memory for Odors. *Journal of Experimental Psychology: Learning Memory and Cognition, 26*(2), 411–422. Retrieved from <https://www.scopus.com/inward/record.url?eid=2-s2.0-0034154649&partnerID=40&md5=9e4b81a094221b6587a9ae6e8da0efbb>
- Roediger III, H. L., & DeSoto, K. A. (2014). Confidence and memory: assessing positive and negative correlations. *Memory (Hove, England), 22*(1), 76–91. <https://doi.org/10.1080/09658211.2013.795974>
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, {&} Computers, 31*(1), 137–149. <https://doi.org/10.3758/BF03207704>
- Stevenson, R. J., & Mahmut, M. K. (2013a). Familiarity influences odor memory stability.

Psychonomic Bulletin & Review, 20(4), 754–9. <https://doi.org/10.3758/s13423-013-0380-9>

Stevenson, R. J., & Mahmut, M. K. (2013b). The accessibility of semantic knowledge for odours that can and cannot be named. *Q J Exp Psychol (Hove)*, 66(7), 1414–1431. <https://doi.org/10.1080/17470218.2012.753097>

Valentin, D., Dacremont, C., & Cayeux, I. (2011). Does short-term odour memory increase with expertise? An experimental study with perfumers, flavourists, trained panellists and novices. *Flavour and Fragrance Journal*, 26(6), 408–415. <https://doi.org/10.1002/ffj.2069>

Wahlheim, C. N., & Jacoby, L. L. (2011). Experience with proactive interference diminishes its effects: mechanisms of change. *Memory & Cognition*, 39(2), 185–195. <https://doi.org/10.3758/s13421-010-0017-4>

Walk, H. A., & Johns, E. E. (1984). Interference and facilitation in short-term memory for odors. *Perception & Psychophysics*, 36(6), 508–14. <https://doi.org/10.3758/BF03207510>

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

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

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>

Appendix 1

Odour	Experiment	Fam.	Int.	Pleas.	Irr.	Verb.
Boiler Room	1	3.04	5.86	2.42	5.42	1.22
Man-o'-War	1	3.14	6.10	1.86	5.73	1.20
Racing Car	1	3.38	6.32	1.88	5.42	1.45

Ginger	1	3.39	5.22	3.10	4.39	1.66
Coal Pit	1	3.49	6.49	1.67	5.96	1.26
Blue Cheese	1	3.51	5.66	1.89	5.60	1.66
Farmyard	1	3.86	5.40	2.46	5.04	1.57
Casbah	1	3.86	4.70	3.86	3.88	1.63
Coriander	1	3.87	4.76	3.98	3.87	1.51
Frosty	1	4.00	3.62	4.49	3.55	1.77
Tea Leaf	1	4.02	4.59	3.80	3.96	1.72
Basil	1	4.06	4.43	4.12	3.37	1.64
Toffee Apple	1	4.10	4.60	4.04	3.78	1.63
Clover	1	4.13	4.02	4.54	3.14	1.63
Rhubarb	1	4.22	4.80	3.92	3.98	1.72
Lychee	1	4.38	4.50	4.44	3.33	1.75
Pine/Heather/Peat	1	4.42	4.64	3.92	3.84	1.83
Cinnamon	1	4.54	5.00	4.44	3.52	1.79
Hot Stuff Male	1	4.62	4.26	4.76	3.00	1.75
Black Pepper	1	4.88	5.04	4.24	3.66	1.77
Aniseed Balls	2	5.88	5.40	3.98	3.54	2.61
Pear	2	5.82	5.16	4.40	3.56	2.62
Pear Drops	2	5.66	5.00	4.74	3.08	2.28
Fruit Punch	2	5.45	4.18	5.29	2.49	2.45
Peach Flesh	2	5.33	3.98	5.20	2.55	2.34
Tea Tree Oil	2	5.20	5.14	3.67	3.80	2.20
Liquorice	2	5.14	4.72	3.56	4.12	2.22
Garlic	2	5.10	6.51	2.06	5.51	2.30
Mango & Sweet						
Orange	2	5.02	4.33	5.12	2.80	2.28

Sage	2	4.96	5.24	4.12	3.58	2.28
Chocolate (Mint)	2	4.90	5.48	4.12	3.84	2.24
Soap Suds	2	4.90	4.14	4.37	3.22	2.54
Fish	2	4.74	5.78	1.66	5.66	2.57
Pineapple	2	4.69	4.35	4.53	3.24	2.36
Onion	2	4.66	6.58	2.00	6.02	2.28
Irish Cream	2	4.63	4.31	4.02	3.58	2.10
Rosemary	2	4.53	4.96	3.84	3.82	2.32
Coal Soot	2	4.27	5.58	2.46	4.88	2.17
Rubber	2	4.12	5.80	2.22	5.39	2.15
Train Smoke	2	3.96	6.06	2.18	5.30	2.39

Fam = Familiarity, Int = Intensity, Pleas = Pleasantness, Irr = Irritability, Verb = Verbalisability

Table 1.

The proportion of correct rejections, and the confidence judgement resolution, for low verbalisability olfactory negative probes. Standard error of the mean is presented in parentheses.

	Non-Recent Negative	Recent Negative
Correct Rejections	.65 (.04)	.64 (.03)
Confidence Resolution	.20 (.12)	.05 (.12)

Table 2.

Comparison of normative data (Moss et al., 2016) for olfactory stimuli in Experiment 1 and 2 (p values are presented in parentheses).

	Experiment 1	Experiment 2	t-test	$B_{N(0, 1)}$
Verbalisability	1.61	2.34	-13.37 (< .001)	-
Familiarity	3.33	4.24	-6.48 (< .001)	-
Intensity	5.00	5.14	-0.55 (.585)	0.27
Pleasantness	3.49	3.68	-0.54 (.601)	0.38
Irritability	4.22	4.00	0.70 (.491)	0.38

Table 3.

The proportion of correct rejections, and the confidence judgement resolution, for high verbalisability olfactory negative probes. Standard error of the mean is presented in parentheses.

	Non-Recent Negative	Recent Negative
Correct Rejections	.70 (.03)	.60 (.05)
Confidence Resolution	.33 (.07)	-.11 (.12)

Table 4.

The proportion of correct rejections, and the confidence judgement resolution, for verbal stimuli negative probes. Standard error of the mean is presented in parentheses.

	Non-Recent Negative	Recent Negative
Correct Rejections	.90 (.02)	.84 (.03)
Confidence Resolution	.14 (.17)	-.33 (.22)

Table 5.

The proportion of correct rejections, and the confidence judgement resolution, for face stimuli negative probes. Standard error of the mean is presented in parentheses.

	Non-Recent Negative	Recent Negative
Correct Rejections	.80 (.03)	.70 (.03)
Confidence Resolution	.13 (.13)	.03 (.10)

Figure 1. Schematic diagram of the recent probes task. The negative probe is taken from the immediately preceding trial in trial 3 (recent), and from three trials previously in trial 4 (non-recent).

Figure 2. Serial position functions for (a) hits, and (b) confidence judgements for hits, for olfactory stimuli used in Experiment 1. Error bars denote mean standard error.

Figure 3. Serial position functions for (a) hits, and (b) confidence judgements for hits, for olfactory stimuli used in Experiment 2. Error bars denote mean standard error.

Figure 4. Serial position functions for (a) hits, and (b) confidence judgements for hits, for verbal stimuli used in Experiment 3. Error bars denote mean standard error.

Figure 5. Serial position functions for (a) hits, and (b) confidence judgements for hits, for facial stimuli used in Experiment 4. Error bars denote mean standard error.