Psychonomic Bulletin & Review
https://doi.org/10.3758/s13423-018-1521-y

BRIEF REPORT

The effect of perceptual information on output interference

Sharon Chen[1] · Kenneth J. Malmberg[2] · Melissa Prince[3] · Shantai Peckoo[1] · Amy H. Criss[1]

© Psychonomic Society, Inc. 2018

Abstract

Retrieval from episodic memory has consequences (Malmberg, Lehman, Annis, Criss, & Shiffrin, *The Psychology of Learning and Motivation,* 61; 285–313, 2014). In some cases, the consequences are beneficial, as in the improvement in memory for items that were already retrieved (Izawa, 1970, *Journal of Experimental Psychology, 83*(2, Pt.1), 340–344; Izawa, *Journal of Experimental Psychology, 89*(1): 10–21, 1971; Roediger & Karpicke, *Psychological Science*, *17*(3), 249–255, 2006). In other cases, the consequences are negative, as in the case of output interference (OI; Wickens, Borne, & Allen, *Journal of Verbal Learning and Verbal Behavior, 2*, 440–445, 1963). OI is the decrease in accuracy in episodic memory with increasing test trials. A release from OI is observed when accuracy rebounds following a switch in the category of item being tested (Criss, Salomão, Malmberg, Aue, Kilic, & Claridge, *Quarterly Journal of Experimental Psychology, 64*(4): 316–326, 2018; Malmberg, Criss, Gangwani, & Shiffrin, *Psychological Science*, *23*(2): 115–119, 2012). In all reports thus far, a release from OI was observed when the conceptual information of stimuli was switched. Here, we evaluate the possibility that changing perceptual information causes a release from OI by presenting items in two perceptual forms (image, audio recording or printed text of the corresponding word) either mixed or blocked at test. A release from OI was observed only for images. We discuss the roles of conceptual and perceptual information in producing OI within the retrieving effectively from memory modeling framework.

Keywords Output interference · Recognition memory · Interference · Perceptual information

Interference prevents memories from being retrieved (e.g., Anderson & Neely, 1996; Raaijmakers & Shiffrin, 1980, 1981; Underwood, 1969). Identifying the sources of interference has long served as a mechanism to understand how information is represented in memory and what information is used to support retrieval. Output interference (OI) is the observation that accuracy decreases across test trials (Ratcliff & Hockley, 1980; Tulving & Arbuckle, 1963, 1966). OI is robust occurring at short and long delays, with and without feedback, and when study–test lag is controlled (Criss, Malmberg, & Shiffrin, 2011). Release from OI is the observation that when the class of test items switches during a testing sequence, memory performance abruptly improves (Criss et al., 2018; Malmberg, Criss, Gangwani, & Shiffrin, 2012; Watkins & Watkins, 1975). For example, Malmberg et al. (2012) presented words from two conceptual categories (e.g., countries vs. professions) randomly intermixed during study. Recognition memory of the words was tested in one of two conditions, either randomly intermixed or blocked by category. OI observed in both conditions. When test was blocked, accuracy increased at the switch point and was near the level of accuracy during the first test bin. Then, performance declined across the second half of testing. This suggests that conceptual information is a critical determinant of interference.

The role of perceptual information in producing OI is unknown, although there is substantial evidence that it is an important component of recognition memory. For example, the format of the stimulus determines overall accuracy. Images are typically remembered with higher accuracy than words are (Nelson, 1979; Paivio, 1971). Visually presented words are better remembered than auditorily presented words (Cleary & Greene, 2002; Gallo, McDermott, Percer, & Roediger, 2001; Smith, Hunt, & Gallagher, 2008). When a test item is perceptually similar to a studied item, fluency increases and memory improves compared with when the format changes between study and test (modality-match effect: Kirsner, 1974; Mulligan & Osborn, 2009; Mulligan, Besken,

✉ Sharon Chen
ychen117@syr.edu

1 Syracuse University, Syracuse, NY, USA

2 University of South Florida, Tampa, FL, USA

3 Flinders University, Adelaide, Australia

Published online: 17 August 2018

& Peterson, 2010; Parks, 2013; Richardson-Klavehn & Bjork, 1988; sensory matching effect: Snodgrass & Hirshman, 1994; Snodgrass, Hirshman, & Fan, 1996), and the word-frequency effect is eliminated when the perceptual aspects of stimuli mismatch (Criss & Malmberg, 2008).

We investigated the role of perceptual information in producing OI by isolating the perceptual format of the stimuli. Each of four experiments manipulated the perceptual form of a concept by presenting an item in the same or different perceptual form at study and at test. Testing was either blocked by perceptual form or intermixed. We expect to see OI in all cases and asked whether changes in perceptual format cause release from OI in the blocked conditions. If only conceptual information contributes to OI, no release should be observed. If both conceptual and perceptual information contribute to OI, release from OI should be observed.

Experiments 1, 2, and 3

General method

Participants

Students from Syracuse University were randomly assigned to condition and received partial course credit.

Materials

Every item was prepared in three perceptual forms—a printed word, an audio recording of the word, and an image, with each experiment including two of the three forms. Images and printed words (labels of the images) were drawn from 240 black-and-white line drawings of Snodgrass and Vanderwart (1980). Audio recordings of each image label were generated by Google translate and recorded into separate.wav files. Every audio file was adjusted so that it has a zero-second delay prior to the sound, its pronunciation was clear and the duration of the word was 1.1 seconds. Sixteen audio files were ambiguous and those stimuli were excluded when one of the two formats in the experiment was audio.

Design and procedure

Each experiment was a 2 × 2 between-subjects design (see Fig. 1). Stimuli from the two perceptual classes were randomly intermixed at study. Test lists were either *blocked* by perceptual form—all stimuli in one perceptual form were presented before all stimuli of the other form—or stimuli were randomly *mixed* without regard to perceptual form. All stimuli in a list were presented in the same (*matched*) or different (*mismatched*) perceptual form at study and test.

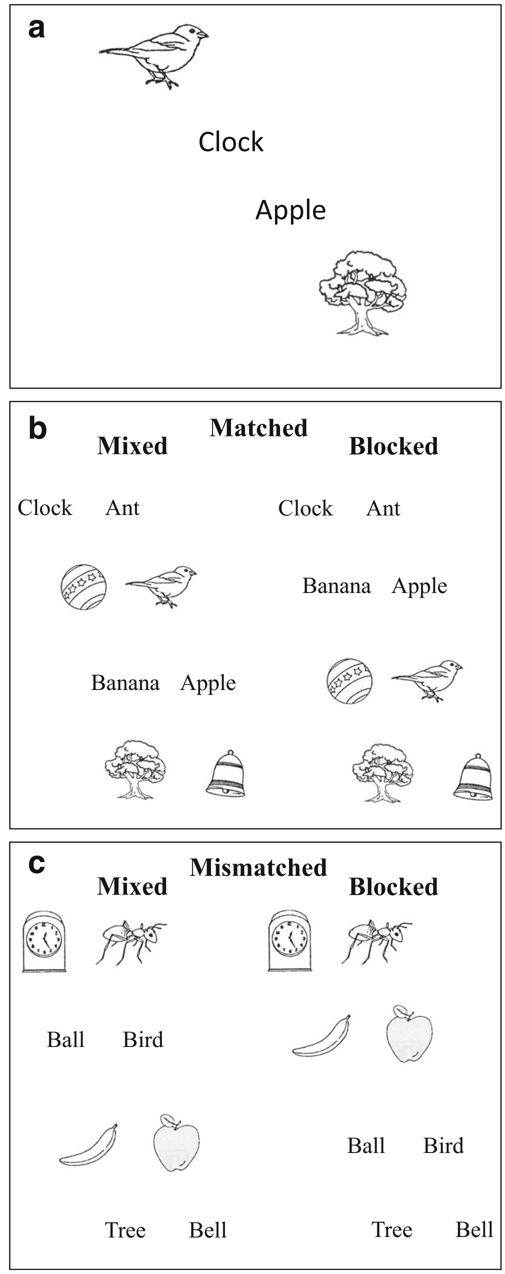

Fig. 1 **a** Example study list for Experiment 1, where each row is a separate study trial. **b** Example test lists in the matched conditions. **c** Example test lists in mismatched conditions. In both **b** and **c**, the left shows intermixed testing and the right shows blocked testing

The study list consisted of 120 trials in Experiment 1 (image, word) and 112 trials in Experiments 2 (audio, word) and 3 (audio, image), half in each perceptual form. Before every stimulus appeared on the center of the screen, participants received a prompt (+) for 0.5 seconds. Then, the stimulus was presented for 1.1 seconds. After the stimulus was displayed, participants answered the question "Do you think you will remember that item?" on a 3-point scale (1 = *definitely no*, 2 = *probably yes*, and 3 = *definitely yes*) by clicking a labeled button. This encoding task is not of theoretical interest for this manuscript, and the data were not analyzed.

After the study phase, participants completed a distractor task requiring them to sum 25 single digits and type the sum in a text box, after which they received feedback. Each digit was presented for 3 seconds with a 0.5 second interstimulus interval (ISI). Next, participants were presented with a series of two-alternative forced-choice test trials (120 in Experiment 1, 112 in Experiments 2 and 3), half in each perceptual format. Immediately following a 0.5 second prompt (+), the two test choices were presented consecutively with position of the studied target and unstudied foil determined randomly on each trial. Consecutive presentation was necessary due to the audio stimuli. Each stimulus presentation was 1.1 seconds in duration and with a 0.5 second ISI. Participants were instructed to select the target by clicking the appropriate button. Buttons were disabled until both stimuli had been shown. The design included full randomization of stimuli to conditions, order of stimuli, and order of conditions.

Data analysis

Forced-choice testing allows us to measure recognition accuracy while minimizing old–new response bias (Grider & Malmberg, 2008; Lockhart & Murdock, 1970). To evaluate OI, we divided the test trials into eight bins of equal size and computed the percentage correct within each bin (note that the pattern of data does not depend on the size of the test bin). In the blocked conditions, perceptual form of test items changed between Bins 4 and 5.

To evaluate OI, we calculated the slope of accuracy across test bin for every participant, and conducted one-sample *t* tests comparing each slope to zero. OI is present if slope is negative. To evaluate whether there is a release from OI when the perceptual form of test items changed, we conducted a 2 (bin4/bin5) × 2 (mixed/blocked) ANOVA (following Criss et al., 2018; Malmberg et al., 2012). Release from OI is present if OI was detected and accuracy in Bin 5 is higher than accuracy in Bin 4 in the blocked condition.

We report Bayes factors (BF), which provide a continuous measure of relative evidence in favor of a hypothesis and are sensitive to sample size and power (e.g., Rouder, Speckman, Sun, Morey, & Iverson, 2009; Wagenmakers, 2007). A BF is the ratio of evidence for one model (the null model in this case) compared with an alternative model (model with an effect in this case). $BF_{01} > 1$ indicates evidence for null model (see Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010). $BF_{01} < 1$ indicates evidence in favor of the alternative model. We do not draw arbitrary labels indicating that any value is "significant" or not (see Etz & Vandekerckhove, 2017; Morey, 2015). Analyses were conducted in JASP (Love et al., 2015) with the default Cauchy prior with width 0.707 (Morey, Rouder, Pratte, & Speckman, 2011; Rouder et al., 2009), which roughly translates as assuming that the effect size is likely to be between ±.70, a relatively large effect size for psychology.

Experiment 1: Words and images

Method

Images and printed words were randomly intermixed during study. We recruited 226 participants, who were randomly assigned to four conditions. Two participants did not complete the experiment, leaving 56 participants assigned to each condition: matched-mixed, mismatched-mixed, matched-blocked, and mismatched-blocked.

Results

As evidenced by negative slopes that differed from zero (see Table 1 and Fig. 2), OI was observed in three of four conditions: matched-mixed, $t(55) = -4.710$, $p < .001$, $BF_{01} < 0.001$; mismatched-mixed, $t(55) = -4.128$, $p < .001$, $BF_{01} = 0.006$; and matched-blocked, $t(55) = -3.418$, $p = .001$, $BF_{01} = 0.043$. There is insufficient evidence regarding OI in the mismatched-blocked condition: $t(55) = -1.865$, $p = .067$, $BF_{01} = 1.361$. To evaluate release from OI when perceptual form changed, we compared accuracy of Bin 4 and Bin 5. There was no main effect of bin, $F(1, 222) = .383$, $p = .537$, $BF_{01} = 8.094$, no main effect of mixed/blocked testing, $F(1, 222) = .015$, $p = .902$, $BF_{01} = 5.865$, and no interaction, $F(1, 222) = .945$, $p = .332$, $BF_{01} = 4.35$. That is, there was no release from OI.

Experiment 2: Words and audio

Method

Printed words and audio recordings were randomly intermixed during study. We recruited 201 participants, who were randomly assigned to four conditions. Two participants did not complete the experiment and two participant numbers were used twice (eliminating another four participants) resulting in 49 participants assigned to matched-mixed, 50 to the mismatched-mixed, 49 to the matched-blocked, and 47 to the mismatched-blocked conditions.

Results

As evidenced by negative slopes that differed from zero (see Table 1 and Fig. 2), OI was observed in three of four

Table 1 Mean slope of accuracy across test bin in each condition and each experiment (one standard error of the mean is in parenthesis)

	Experiment 1: Words and images		Experiment 2: Words and audio		Experiment 3: Images and audio	
	Matched	Mismatched	Matched	Mismatched	Matched	Mismatched
Mixed test	−0.009 (0.002)	−0.011 (0.003)	−0.007 (0.002)	−0.016 (0.002)	−0.004 (0.002)	−0.008 (0.003)
Blocked test	−0.010 (0.003)	−0.006 (0.003)	−0.012 (0.003)	−0.005 (0.003)	−0.001 (0.003)	−0.007 (0.003)

conditions: matched-mixed, $t(48) = -3.640$, $p < .001$, $BF_{01} = 0.024$; mismatched-mixed, $t(49) = -4.095$, $p < .001$, $BF_{01} = 0.007$; and matched-blocked, $t(48) = -7.109$, $p < .001$, $BF_{01} < 0.001$. As in Experiment 1, there was insufficient evidence regarding OI in the mismatched-blocked condition, $t(46) = -1.405$, $p = .167$, $BF_{01} = 2.521$. To evaluate whether there is a release from OI, we compared accuracy of Bin 4 and Bin 5. There was no main effect of bin, $F(1, 193) = 1.265$, $p = .262$, $BF_{01} = 4.962$, no effect of list type, $F(1, 193) = 1.418$, $p = .235$, $BF_{01} = 5.336$, and no interaction, $F(1, 193) = .227$, $p = .634$, $BF_{01} = 2.34$. We found no evidence for release from OI, when perceptual form changes, replicating Experiment 1.

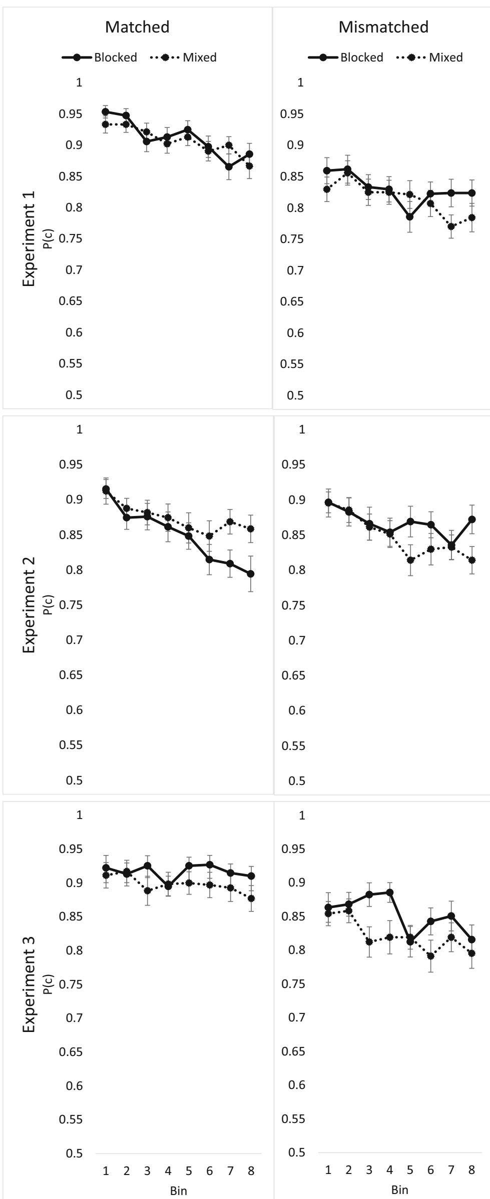

Fig. 2 Means percentage correct (P (c)) across test bin. Experiment 1 contained words and images, Experiment 2 contained words and audio, and Experiment 3 included images and audio presentation of stimuli. Perceptual form of test items was switched between Bin 4 and Bin 5. Error bars are one standard error of the mean above and below

Experiment 3: Audio and images

Method

Audio recordings and images were randomly intermixed during study. We recruited 199 participants, who were randomly assigned to the four conditions. Two participants did not complete the experiment, and two participant numbers were used twice, resulting in 50 participants assigned to matched-mixed, 51 to the mismatched-mixed, 47 to the matched-blocked, and 45 to the mismatched-blocked conditions.

Results

As indicated by negative slopes that differed from zero (see Table 1 and Fig. 2), OI was observed in mismatched conditions. mismatched-blocked: $t(44) = -2.372$, $p = .022$, $BF_{01} = 0.502$; mismatched-mixed: $t(50) = -2.786$, $p = 0.008$, $BF_{01} = 0.210$. In the matched conditions, the slopes did not differ from zero, matched-blocked: $t(46) = -0.171$, $p = .865$, $BF_{01} = 6.228$; matched-mixed: $t(49) = -1.693$, $p = .097$, $BF_{01} = 1.725$. To evaluate release from OI we compared accuracy of Bin 4 and Bin 5. We observed no main effect of bin, $F(1, 191) = .875$, $p = .351$, $BF_{01} = 6.234$, no main effect of list type, $F(1, 191) = 1.779$, $p = .184$, $BF_{01} = 2.615$, and no interaction, $F(1, 191) = 1.007$, $p = .317$, $BF_{01} = 4.14$. We found no evidence that changing perceptual form causes release from OI, replicating Experiments 1 and 2.

Discussion

OI was observed in eight out of 12 conditions. In no case was a release from OI observed when perceptual form of test items changed. This contrasts with results of prior experiments where a release from OI was observed when category, or conceptual information, of the test items changed (Criss et al., 2018; Malmberg et al., 2012) and suggests that perceptual information does not contribute to OI. However, OI was not observed in several conditions, making the results difficult to interpret.

Why was OI not observed in every condition? One possibility is the high accuracy in these experiments. Kilic, Criss, Malmberg, and Shiffrin (2017) showed that the magnitude of OI is dependent on overall accuracy for both model predictions and human behavior. The intuition is straightforward considering Criss et al. (2011), which attributed OI to item information encoded at test. If accuracy is already high then there is a limited amount of additional encoding that can take place during test. If encoding during test is limited, then so too will be OI. To further investigate this hypothesis, we evaluated overall accuracy as a function of perceptual form (aggregating data from three experiments). Inspection of Table 2 shows that accuracy for images was at ceiling when the study and test format matched. Further analysis demonstrated a modality-match effect for images but not the other perceptual forms. We conducted a 2 (matched/mismatched) × 2 (mixed/blocked) ANOVA on percentage correct for each of the perceptual forms (see Table 2). Accuracy of matched conditions was better than mismatched conditions for images, $F(1, 413) = 175.4$, $p < .001$, $BF_{01} < 0.001$, but not in words, $F(1, 415) = 1.729$, $p = .189$, $BF_{01} = 3.944$, or audio stimuli, $F(1, 384) = 0.185$, $p = .667$, $BF_{01} = 7.922$. There was no main effect of mixed/blocked, words: $F(1, 415) = 0.063$, $p = .801$, $BF_{01} = 8.950$; images: $F(1, 413) = 2.490$, $p = .115$, $BF_{01} = 3.545$; audio: $F(1, 384) = 1.101$, $p = .295$, $BF_{01} = 5.243$, and no interaction between matched/mismatched and mixed/blocked, words: $F(1, 415) = 3.646$, $p = .057$, $BF_{01} = 1.28$; images: $F(1, 413) = 0.206$, $p = .650$, $BF_{01} = 5.62$; audio: $F(1, 384) = 2.124$, $p = .146$, $BF_{01} = 2.24$.

In short, we observed a modality-match effect for images and high accuracy for images in the matched testing conditions. This suggests that OI was not observed due to ceiling effects (and release from OI cannot be measured in the absence of OI). In addition, these results suggest that, at test, perceptual form is critically important for images, but not for words or audio. Based on these findings, we propose that if accuracy were at a reasonable level, then we should see both OI and release from OI for images. The similarity in accuracy for the matched and mismatched conditions for words and audio stimuli suggests that perceptual form is not important for these stimuli, consistent with the finding of no release from OI.

We conducted two additional experiments to evaluate this hypothesis. The primary difference between these experiments and the earlier ones is the use of a single-item recognition (SIR) task for testing rather than a two-alternative forced-choice task. There are several reports of OI in SIR in the literature, all of which show that hit rates decrease across test trial and false-alarm rates tend to be flat on average—sometimes slightly increasing, sometimes slightly decreasing, and often unchanged across test bin (e.g., Criss et al., 2011; Kilic et al., 2017; Koop, Criss, & Malmberg, 2015). In the following experiments, we include only the most informative conditions—those where performance was near ceiling for images (the matched conditions) and release from OI could be observed (the blocked testing).

Experiments 4a and 4b

General method

Participants and materials

Participants were selected from the same pool and materials were identical to Experiments 1–3.

Design and procedure

Participants were randomly assigned to Experiment 4a (words and images) or Experiment 4b (images and audio), and stimuli were randomly assigned to condition for each participant. Stimuli were presented in a matched form in study and test, and test lists were blocked by perceptual form. Order of block was random and stimuli within a block were randomized. At test, participants were presented with an equal number of studied targets and unstudied foils randomly selected. Participants responded to the question "Have you studied this item?" by pressing a button, YES or NO, displayed below the test stimulus. All remaining details were identical to Experiments 1–3.

Table 2 Mean accuracy in each condition by perceptual form (aggregating Experiments 1, 2, and 3). One standard error of the mean is presented in parenthesis

	Words		Images		Audio	
	Matched	Mismatched	Matched	Mismatched	Matched	Mismatched
Mixed test	0.87 (0.01)	0.84 (0.01)	0.93 (0.02)	0.79 (0.01)	0.87 (0.01)	0.85 (0.01)
Blocked test	0.86 (0.01)	0.86 (0.01)	0.95 (0.01)	0.80 (0.01)	0.87 (0.01)	0.88 (0.01)

Data analysis plan

Test trials were divided into eight bins of equal size (perceptual form of the test items changed between Bins 4 and 5), and we computed hit rates and false-alarm rates within each bin. We evaluate OI by comparing slopes to zero with one-sample *t* tests and evaluate release from OI by comparing Bins 4 and 5. In a blocked design, participants responded to one form of stimuli in Bins 1–4 and for the other for Bins 5–8. Therefore, we compute slopes separately for Bins 1–4 and Bins 5–8.

Experiment 4a: Words and images

Method

Images and printed words were randomly intermixed during study. We recruited 73 participants. One participant did not complete the experiment, and two participant numbers were used twice, resulting in 68 participants. Thirty-six were tested on word stimuli first, and 32 were tested on images first.

Results

OI was observed in every case and in the expected form—a decrease in hit rate and relatively flat false-alarm rate (see Table 3). Slopes of the hit rates were negative and differed from zero, words—Bins 1–4: $t(35) = -2.624$, $p = .013$, $BF_{01} = 0.291$; words—Bins 5–8: $t(31) = -2.432$, $p = .021$, $BF_{01} = 0.423$; images—Bins 1–4: $t(31) = -4.667$, $p < .001$, $BF_{01} = 0.002$; images—Bins 5–8: $t(35) = -4.489$, $p < .001$, $BF_{01} = 0.003$. The slope of the false alarm rates did not differ from zero, words—Bins 1–4: $t(35) = 1.433$, $p = .161$, $BF_{01} = 2.192$; words—Bins 5–8: $t(31) = 1.360$, $p = .184$, $BF_{01} = 2.292$; images—Bins 1–4: $t(31) = -1.465$, $p = .153$, $BF_{01} = 2.012$; images—Bins 5–8: $t(35) = -0.396$, $p = .695$, $BF_{01} = 5.191$.

To evaluate whether there is a release from OI when perceptual form changed, we compared the hit rates in Bin 4 and Bin 5 (see Fig. 3). A release from OI was observed in hit rate of images, $t(66) = -4.040$, $p < .001$, $BF_{01} = 0.006$, but not in hit rate of words, $t(66) = 1.676$, $p = .099$, $BF_{01} = 1.324$. False-alarm rates did not differ from Bin 4 to Bin 5, words, $t(66) = -0.085$, $p = .932$, $BF_{01} = 3.998$; images, $t(66) = 0.864$, $p = .391$, $BF_{01} = 2.917$.

Experiment 4b: Images and audio

Method

Images and audio recordings were randomly intermixed during study. We recruited 70 participants. One participant number was used twice, resulting in 68 participants. Of 68 participants, we have 34 subjects who were tested on audio then images, and 34 tested on images then audio.

Results

OI was observed in hit rates for every condition (see Table 3). All slopes were negative and differed from zero, images—Bins 1–4: $t(33) = -3.881$, $p < .001$, $BF_{01} = 0.016$; images—Bins 5–8: $t(33) = -5.774$, $p < .001$, $BF_{01} < 0.001$; audio—Bins 1–4: $t(33) = -3.613$, $p < .001$, $BF_{01} = 0.031$; audio—Bins 5–8: $t(33) = -2.913$, $p = .006$, $BF_{01} = 0.158$. The slope of the false-alarm rates did not differ from zero, images—Bins 1–4: $t(33) = 0.639$, $p = .527$, $BF_{01} = 4.501$; images—Bins 5–8: $t(33) = 1.565$, $p = .127$, $BF_{01} = 1.803$; audio—Bins 1–4: $t(33) = 1.237$, $p = .225$, $BF_{01} = 2.705$; audio—Bins 5–8: $t(33) = 1.207$, $p = .236$, $BF_{01} = 2.794$.

To evaluate release from OI when perceptual form changed, we compared the hit rates in Bin 4 and Bin 5 (see Fig. 3). A release from OI was observed for images, $t(66) = -3.087$, $p = .003$, $BF_{01} = 0.080$, but not audio recordings, $t(66) = -0.309$, $p = .759$, $BF_{01} = 3.856$. False-alarm rates of audio stimuli did not differ between Bins 4 and 5, $t(66) = -0.225$, $p = .822$, $BF_{01} = 3.930$. The false-alarm rate of images in Bin 5 was lower than in Bin 4, $t(66) = 2.496$, $p = .015$, $BF_{01} = 0.300$.

Discussion

When accuracy was not at ceiling, we observed OI in for all perceptual forms. Release from OI was only present for images. This is consistent with the observation the large benefit for matched over mismatched testing for images in Experiments 1–3. Together, these data provide evidence that

Table 3 Mean slope of hit rate and false alarm rate as a function of Test Bins 1–4, and Bins 5–8, in words and images from Experiment 4a and, images and audio from Experiment 4b. One standard error of the mean is in parenthesis

	Experiment 4a: Words		Experiment 4a: Images		Experiment 4b: Images		Experiment 4b: Audio	
	Bins 1–4	Bins 5–8	Bins 1–4	Bins 5–8	Bins 1–4	Bins 5–8	Bins 1–4	Bins 5–8
Hit rate	−0.02 (0.004)	−0.03 (0.001)	−0.05 (0.003)	−0.06 (0.004)	−0.04 (0.004)	−0.06 (0.003)	−0.04 (0.002)	−0.03 (0.002)
False-alarm rate	0.01 (0.001)	0.01 (0.001)	−0.01 (0.001)	0.00 (0.003)	0.01 (0.001)	0.01 (0.001)	0.01 (0.001)	0.01 (0.002)

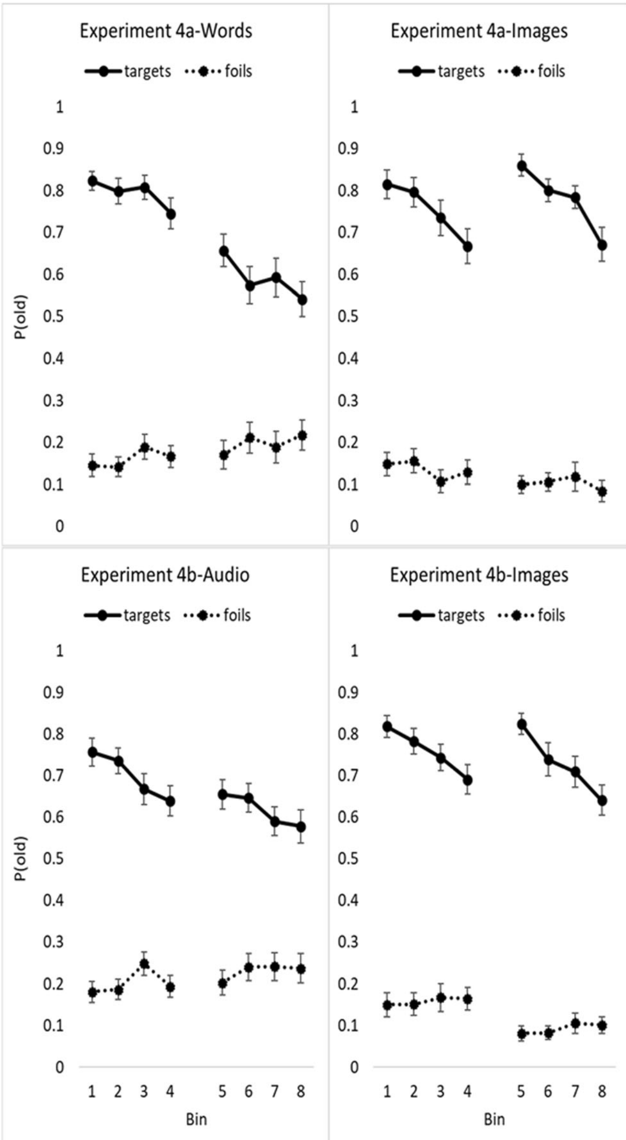

Fig. 3 Probability of responding old (P (old)) for targets (hit rate) and foils (false-alarm rate) across test bin. Perceptual form of test items was switched between Bin 4 and Bin 5. The gap between Bin 4 and Bin 5 is a reminder that participants contributing to Bins 1–4 are different from those in Bins 5–8. Error bars are one standard error of the mean above and below

perceptual information contributes to episodic memory generally and to OI for images.

General discussion

We investigated the role of perceptual information in OI. We defined perceptual form in terms of the form of a stimulus used to represent a single concept (i.e., black-and-white line drawing, text of label, or audio recording of the label for the same concept); hence, we held conceptual information constant while varying perceptual information. We manipulated whether the perceptual form of the stimulus matched or mismatched between study and test, and we manipulated whether the test list was blocked by presentation form or intermixed. Memory for word and audio stimuli suffered from OI but did not benefit from a change in perceptual form. Further, memory was not affected by whether such stimuli were studied and tested in the same format. On the other hand, images were affected by perceptual form. Memory for images suffered from OI and benefited from a release from OI when stimulus type changed at test. Further, accuracy for images was greatly enhanced when the perceptual form was identical between study and test (i.e., the modality-match effect; Mulligan & Osborn, 2009; Mulligan et al., 2010).

Why might images alone be influenced by perceptual form? One likely explanation is that an image is not an abstract concept, like “bird,” but instead represents a very specific exemplar of a bird (e.g., the one pictured in Fig. 1). While images are reliably given the labels, which we used as stimuli in the other perceptual formats (e.g., see Snodgrass & Vanderwart, 1980), any of a number of images may map on to that same label or may be evoked in the mind’s eye when given the label. In other words, there are additional features associated with the image stimuli that support memory. If these features are encoded during study, then the absence of those features during test may alter memory accuracy. Words and audio recordings have many fewer external features that support memory. In contrast, word and audio stimuli represent the same abstract concept as the image (“bird”) but lack many of the additional distinct features associated with images. We suggest that it is those rich features found in the images that are enabling release from OI. One could imagine rich audio recordings (e.g., the word *bird* with a bird sound) or rich text (printed in a unique font, i.e., Reder, Donavos, & Erickson, 2002) that might also provide the basis for a release from OI. In the absence of rich information, audio and printed presentation provide insufficient information for memory to be selectively influenced by perceptual form.

The rich set of OI data spanning several papers is best accounted for within the computational framework of retrieving effectively from memory (REM; Shiffrin & Steyvers, 1997; see Kilic et al., 2017, for a comprehensive technical discussion of OI in REM, and Criss & Koop, 2015, for a conceptual discussion). Criss et al. (2011) extended the model to account for OI by assuming that the cue used to probe memory was encoded during retrieval. When a test stimulus was remembered as old, the best matching memory trace was updated with information about the test stimulus. When a test stimulus was believed to be new, a new memory trace was encoded. These specific assumptions are necessary to capture detailed patterns of OI (see Criss et al., 2011; Kilic et al., 2017, for more information). Within REM, OI is due to interference from stimulus information encoded at test. Interference builds up to the degree that the test stimulus shares features with previously encoded traces. Repetition of features shared by a category of items (such as a list of country names) increases interference with subsequently tested items from the same category. A change in category (such as country names to profession names) leads to a temporary reduction in interference due to encoding of fewer relevant features. In other words, release from OI is due to salient difference in stimulus information used to cue memory. Consistent with the results here (and see Kilic et al., 2017, for full details), the better the initial stimulus is encoded, the less information is available to be updated during test.

REM does not explicitly address perceptual or conceptual information but could potentially account for the current findings under the assumption that images share features (such as shading, nonletter lines and dots, etc.) that are shared by other images, but not shared with words in printed or audio format. Interference from these features builds up as they are encoded during test. This is similar to our explanation for release from OI observed for faces (when studied with words; Criss et al., 2018). What is perhaps surprising is that the rebound appears to be complete, or, performance returns to the same level as the first test bin. This means, at least within REM, the interference between images and other perceptual forms is minimal. Words printed or spoken share largely the same set of features and therefore share interference. There are, of course, a few features that are isolated to printed words (e.g., the font) or the audio recordings (speakers voice), but these are small enough in number or insufficiently rich in content that they have little impact on memory. Next, we consider potential alternative explanations.

First, consider a context-drift explanation. On one hand, context might drift due to the passing of time or slowly changing items, a global context. This would cause any test cue to be remembered less well as trials (or time) pass, reducing performance. While reasonable, this account fails to explain release from OI (Criss et al., 2018; Malmberg et al., 2012). On the other hand, context drift may be due to external factors related to the stimulus, such as perceptual form. Perceptual form is often assumed to be a context-based manipulation (e.g., Hirshman, Passanante, & Arndt, 1999; McGeoch, 1932; Murnane & Phelps, 1995), especially when the perceptual form is the same for a large number items, which is another characteristic that defines context (Lehman & Malmberg, 2013; Malmberg & Shiffrin, 2005). In this case, it is unclear why this local context should matter only for images, but not for other stimuli, or why image was used as a context cue, but printed or audio format was not. It is possible to attribute this to the "distinctiveness" of images, but it is equally reasonable to claim that audio recordings are distinctive (heard, not read, in a robotic voice). In other words, absent a concrete theory-based definition of distinctiveness, this construct carries little explanatory power.

A second potential account of OI is the attention hypothesis (e.g., Criss, et al., 2018; Watkins & Watkins, 1975; Wickens, 1970), according to which attention decreases with increasing test trials, and the observed decrease in accuracy is due to lack of effort, attention, or motivation. A release from OI may be observed when changing stimulus classes, conceptually or perceptually, boosts attention and, therefore, accuracy. However, when the attention hypothesis has been directly tested, it has been disconfirmed (e.g., Criss et al., 2018), and the current data do not provide any support for the attention hypothesis. Relatedly, retention interval (or study–test lag) is a plausible candidate. Criss et al. (2011) directly tested this by presenting items in the same order at study and test, controlling for study–test lag and retention interval and found OI of the same slope and magnitude as when items are randomly ordered.

Last, consider a cue-overload hypothesis (e.g., Watkins & Watkins, 1975), which attributes the degree of interference to the greater number of associations (e.g., items) connected to any specific cue (e.g., audio). When the cue changes (e.g., audio to images), the overload resets, eliciting more efficient retrieval and a release from interference. The REM model (Criss et al., 2011) incorporates this cue-overload principle at the level of the feature rather than the item. The cue-overload hypothesis itself lacks power to account for the existence of release from OI at switch point in images, and lack of release from OI in other formats without resorting to a feature-based interpretation.

In conclusion, the major finding from this project is that perceptual information contributes to OI, providing that this perceptual information contributes to retrieval process. As Underwood (1969) proposed, memories are composed of multiple attributes that can be used to cue memory. When such cuing occurs, these attributes, conceptual or perceptual, are source of interference in episodic memory.

Acknowledgements S.P. was supported by the Psychology Research in Diversity Enhancement program of the Department of Psychology at Syracuse University. M.P. is now at Flinders University. Authors are listed in order of their degree of involvement in the work, with the most substantial contributors listed first. All data and audio files are available at https://osf.io/znk4f/.

References

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. Bjork & R. A. Bjork (Eds.), Memory (pp. 237–313). San Diego, CA: Academic Press.

Cleary, A. M., & Greene, R. L. (2002). Paradoxical effects of presentation modality on false memory. *Memory, 10*, 55– 61. doi:https://doi.org/10.1080/09658210143000236

Criss, A. H., & Koop, G. J. (2015). Differentiation in episodic memory. In J. Raaijmakers, A. H. Criss, R. Goldstone, R. Nosofsky, & M.

Steyvers (Eds.), Cognitive modeling in perception and memory: A Festschrift for Richard M. Shiffrin (pp. 112–125). New York: Psychology Press.

Criss, A. H., & Malmberg, K. J. (2008). Evidence in support of the elevated attention hypothesis of recognition memory. *Journal of Memory & Language, 59*, 331–345. doi:https://doi.org/10.1016/j.jml.2008.11.001

Criss, A. H., Malmberg, K. J., & Shiffrin, R. M. (2011). Output interference in recognition memory. *Journal of Memory and Language, 64,* 316–326. doi:https://doi.org/10.1016/j.jml.2011.02.003

Criss, A. H., Salomão, C., Malmberg, K. J., Aue, W. A., Kilic, A., & Claridge, M. (2018). Release from output interference in recognition memory: A test of the attention hypothesis. *Quarterly Journal of Experimental Psychology, 64*(4), 316–326. doi:https://doi.org/10.1080/17470218.2017.1310265

Etz, A., & Vandekerckhove, J. (2017). Introduction to Bayesian inference for psychology. *Psychonomic Bulletin & Review, 25*(1), 5–34. doi:https://doi.org/10.3758/s13423-017-1262-3

Gallo, D. A., McDermott, K. B., Percer, J. M., & Roediger, H. L., III. (2001). Modality effects in false recall and false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 339–353. doi:https://doi.org/10.1037/0278-7393.27.2.339

Grider, R. L., & Malmberg, K. J. (2008). Discriminating between changes in bias and changes in accuracy for recognition memory of emotional stimuli. *Memory & Cognition*, *36*(5), 933–946. doi:https://doi.org/10.3758/MC.36.5.933

Hirshman, E., Passanante, A., & Arndt, J. (1999). The effect of midazolam on the modality-match effect in implicit memory. *Cognitive Brain Research*, *7,* 473–549. doi:https://doi.org/10.1016/S0926-6410(98)00046-9

Izawa, C. (1970). Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. *Journal of Experimental Psychology, 83*(2, Pt.1), 340–344. doi:https://doi.org/10.1037/h0028541

Izawa, C. (1971). Massed and spaced practice in paired-associate learning: List versus item distributions. *Journal of Experimental Psychology, 89*(1), 10–21. doi:https://doi.org/10.1037/h0031177

Kilic, A., Criss, A. H., Malmberg, K. J., & Shiffrin, R. M. (2017). Models that allow us to perceive the world more accurately also allow us to remember past events more accurately via differentiation. *Cognitive Psychology*, *92,* 65–86. doi:https://doi.org/10.1016/j.cogpsych.2016.11.005

Kirsner, K. (1974). Modality differences in recognition memory for words and their attributes. *Journal of Experimental Psychology*, *102,* 579–584. doi:https://doi.org/10.1037/h0036112

Koop, G. J., Criss, A. H., & Malmberg, K. J. (2015). The role of mnemonic processes in pure-target and pure-foil recognition memory. *Psychonomic Bulletin & Review, 22*(2), 509–516. doi:https://doi.org/10.3758/s13423-014-0703-5

Lehman, M., & Malmberg, K. J. (2013). A buffer model of encoding and temporal correlations in retrieval. *Psychological Review, 120*(1), 155–189. doi:https://doi.org/10.1037/a0030851

Lockhart, R. S., & Murdock, B. B. (1970). Memory and the theory of signal detection. *Psychological Bulletin*, *74*(2), 100–109. doi:https://doi.org/10.1037/h0029536

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., … Wagenmakers, E. J. (2015). JASP (Version 0.7.1) [Computer software]. Retrieved from https://jasp-stats.org/

Malmberg, K. J., Criss, A. H., Gangwani, T. H., & Shiffrin, R. M. (2012). Overcoming the negative consequences of interference from recognition memory testing. *Psychological Science*, *23*(2), 115–119. doi:https://doi.org/10.1177/0956797611430692

Malmberg, K. J., Lehman, M., Annis, J., Criss, A. H., & Shiffrin, R. M. (2014). Consequences of testing memory. In B. Ross (Ed.), The *psychology of learning and motivation* (Vol. 61, pp. 285–313). San Diego, CA: Academic Press.

Malmberg, K. J., & Shiffrin, R. M. (2005). The "one-shot" hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 322–336. doi:https://doi.org/10.1037/0278-7393.31.2.322

McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychology Review, 39*, 352–370.

Morey, R. (2015). On verbal categories for the interpretation of Bayes factors [Web log post]. Retrieved from http://bayesfactor.blogspot.com.au/2015/01/on-verbal-categories-for-interpretation.html

Morey, R. D., Rouder, J. N., Pratte, M. S., & Speckman, P. L. (2011). Using MCMC chain outputs to efficiently estimate Bayes factors. *Journal of Mathematical Psychology, 55,* 368–378.

Mulligan, N. W., Besken, M., & Peterson, D. (2010). Remember-know and source memory instructions can qualitatively change old-new recognition accuracy: The modality-match effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 36,* 558–566. doi:https://doi.org/10.1037/a0018408

Mulligan, N. W., & Osborn, K. (2009). The modality-match effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 564–571. doi:https://doi.org/10.1037/a0014524

Murnane, K., & Phelps, M. P. (1995). Effects of changes in relative cue strength on context-dependent recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21,* 158–172.

Nelson, D. L. (1979). Remembering pictures and words: Appearance, significance, and name. In L. S. Cermak & F. I. M. Craik (Eds.), Levels of processing in human memory (pp. 45–76). Hillsdale: Erlbaum.

Paivio, A. (1971). Imagery and verbal processes. London: Holt, Rinehart & Winston.

Parks, C. M. (2013). Transfer-appropriate processing in recognition memory: Perceptual and conceptual effects on recognition memory depend on task demand. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(4), 1280–1286. doi:https://doi.org/10.1037/a0030911

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 14, pp. 207–262). New York: Academic Press.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*(2), 93–134. doi:https://doi.org/10.1037/0033-295X.88.2.93

Ratcliff, R., & Hockley, W. E. (1980). Repeated negatives in item recognition: Nonmonotonic lag functions. In R. S. Nickerson (Ed.), Attention and performance VIII. Hillsdale: Erlbaum.

Reder, L., Donavos, D., Erickson, M. (2002). Perceptual match effects in direct tests of memory: the role of contextual fan. *Memory & Cognition*, 30, 2, 312-323. Doi https://doi.org/10.3758/BF03195292

Richardson-Klavehn, A., & Bjork, R. A. (1988). Measures of memory. *Annual Review of Psychology*, *36,* 475–543. doi:https://doi.org/10.1146/annurev.ps.39.020188.002355

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255. doi:https://doi.org/10.1111/j.1467-9280.2006.01693

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16,* 225–237. doi:https://doi.org/10.3758/PBR.16.2.225

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM: Retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4,* 145–166. doi:https://doi.org/10.3758/BF03209391

Smith, R. E., Hunt, R. R., & Gallagher, M. P. (2008). The effect of study modality on false recognition. *Memory & Cognition, 36*, 1439–1449. doi:https://doi.org/10.3758/MC.36.8.1439

Snodgrass, J. G., & Hirshman, E. (1994). Dissociations among implicit and explicit memory tasks: The role of stimulus similarity. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *20,* 150–160.

Snodgrass, J. G., Hirshman, E, & Fan, J. (1996). The sensory match effect in recognition memory: Perceptual fluency or episodic trace? *Memory & Cognition*, *24*(3), 367–383. doi:https://doi.org/10.3758/BF03213300

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: Learning, Memory, and Cognition*, *6*(2), 174–215.

Tulving, E., & Arbuckle, T. (1963). Sources of intratrial interference in immediate recall of paired associates. *Journal of Verbal Learning & Verbal Behavior*, *1,* 321–334. doi:https://doi.org/10.1016/S0022-5371(63)80012-2

Tulving, E., & Arbuckle, T. (1966). Input and output interference in short-term associative memory. *Journal of Experimental Psychology*, *72,* 145–150.

Underwood, B. J. (1969). Attributes of memory. *Psychological Review*, *76,* 559–573.

Wagenmakers, E., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology, 60*(3), 158–189. doi:https://doi.org/10.1016/j.cogpsych.2009.12.001

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, *14*, 779–804. doi: https://doi.org/10.3758/BF03194105

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory, 104,* 442–452. doi:https://doi.org/10.1037/0278-7393.1.4.442

Wickens, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review*, *77,* 1–15. doi:https://doi.org/10.1037/h0028569

Wickens, D. D., Borne, D. G., & Allen, C. K. (1963). Proactive inhibition and item similarity in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 2*, 440–445. doi:https://doi.org/10.1016/S0022-5371(63)80045-6