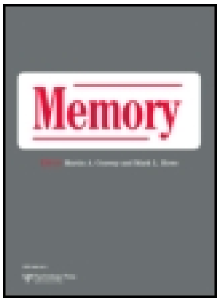

Memory
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pmem20

Interplay of the production and picture superiority effects: A signal detection analysis

Jonathan M. Fawcett [a], Chelsea K. Quinlan [a] & Tracy L. Taylor [a]
[a] Department of Psychology, Dalhousie University, Halifax, NS, Canada

Available online: 26 Jun 2012

MEMORY, 2012, iFirst, 1–12

Psychology Press
Taylor & Francis Group

Interplay of the production and picture superiority effects: A signal detection analysis

Jonathan M. Fawcett, Chelsea K. Quinlan, and Tracy L. Taylor

Department of Psychology, Dalhousie University, Halifax, NS, Canada

Three experiments explored the interaction between the production effect (greater memory for produced compared to non-produced study items) and the picture superiority effect (greater memory for pictures compared to words). Pictures and words were presented in a blocked (E1) or mixed (E2, E3) design, each accompanied by an instruction to silently name (non-produced condition) or quietly mouth (produced condition) the corresponding referent. Memory was then tested for all study items as well as an equal number of foil items using a speeded (E1, E2) or self-paced (E3) yes-no recognition task. Experiments 1, 2, and 3 all revealed a small but reliable production $\times$ stimulus interaction. Production was also found to result in a liberal shift in response bias that could result in the overestimation of the production effect when measured using hits instead of sensitivity. Together our findings suggest that the application of multiple distinctive processes at study produces an especially discriminative memory trace at test, more so than the summation of each process individually.

Keywords: Production effect; Picture superiority effect; Human memory; Cognition.

Distinct memories tend to stand out, rendering them more accessible than their less-distinctive counterparts (e.g., Hunt, 2006). In the laboratory many studies have investigated the contributions of distinctiveness to human memory. One recent paradigm requires participants to silently read or produce a series of study items for which they are subsequently tested (see MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010). Studies using this technique have demonstrated superior recall or recognition of produced relative to silently read (non-produced) items across a range of production tasks, which include quietly saying the item aloud (e.g., Hourihan & MacLeod, 2008), silently mouthing the item (MacLeod et al., 2010, Expt 5), and writing the item on a tablet (MacLeod, 2010). The benefits associated with producing as opposed to silently reading an item have been referred to as the *production effect* (e.g., MacLeod et al., 2010). The notion is that participants retain a ‘‘production trace’’ of the relevant items at study and that these production traces may then be retrieved to guide test performance. In this sense the produced items are thought to be distinctive relative to the non-produced items because whereas the non-produced items are characterised by a lexical trace alone, the produced items are associated with a production trace. Schacter, Israel, and Racine (1999) argued that participants could infer whether an item had been studied based on the availability of its previously encoded distinctive details. Accordingly, if participants are able to access any element of the earlier production trace

Address correspondence to: Jonathan M. Fawcett, MSc, Department of Psychology, Dalhousie University, Halifax, NS, Canada, B3H 4J1. E-mail: jmfawcet@dal.ca

JMF was funded by a Killam Predoctoral Fellowship. CKQ was funded by an NSERC Vanier Canadian Graduate Scholarship and also holds an honorary Killam Predoctoral Fellowship. TLT was funded by an NSERC Operating Grant. The authors would like to thank Thanh Luu and Emily Nichols for their help implementing this project and Dr Colin MacLeod for sharing his thoughts on these experiments.

http://www.psypress.com/memory http://dx.doi.org/10.1080/09658211.2012.693510

for a given item they can infer the item had been studied because distractors are unlikely to have been recently produced (for discussion, see Dodson & Schacter, 2001).[1]

Several findings have supported distinctiveness as an account for the memory advantage observed for produced compared to non-produced items. MacLeod et al. (2010) argued that the production effect occurs only when the produced items are intermixed with non-produced items. This is because the *relative* distinctiveness between the produced and non-produced items is critical for obtaining the effect (although see Fawcett, 2012). Ozubko and MacLeod (2010) further tested the distinctiveness account by presenting participants with two study lists instead of one: (a) a critical mixed list of study words for which half of the words were produced and half of the words were not produced, and (b) a list of distractors for which half of the participants produced all the words and half of the participants did not. The study and distractor lists were then combined and participants were asked to identify the list from which each item had been drawn. The researchers hypothesised that producing the distractors would equate the processing of these items with the processing of the produced study items. Because equating the processing of study and distractor items eliminates the relative distinctiveness of the study items, Ozubko and MacLeod (2010) expected this manipulation to eliminate the production effect. This prediction was supported. When participants silently read the list of distractors there was a production effect for the study items; however, when participants produced the list of distractors there was no production effect for the study items. Because the production effect is eliminated when study and distractor items are both produced, it suggests that the recollection of distinct production details at test is a necessary component for the production effect.

Given that the production effect is attributed to distinctiveness, it raises the question as to whether it would interact with other effects that are also attributed to distinctiveness. The primary goal of our experiment was to address this question by comparing the production effect for pictures and words. Memory for pictures is typically better than memory for words, a result that is referred to as the picture superiority effect (Madigan, 1983; Paivio, 1991). It has been suggested that the unique visual details inherent to pictures, which are not present in words, provides a distinctiveness cue that helps participants discriminate study items from distractor items (see Lloyd & Miller, 2011). Similar to the distinctiveness account of the production effect, the notion is that participants encode a rich visual trace of the studied pictures resulting in many individual features that could be used to guide test performance. In this sense pictures are distinctive relative to words because whereas words are characterised by a lexical trace alone, memory for pictures includes additional visual information (e.g., the shape or colour of an item).

The distinctiveness interpretation of the production effect and the picture superiority effect suggests a common retrieval process operating on separate representational dimensions.[2] Produced items are thought to be distinctive because the act of production associates a production trace with the representation of the study item in memory and it is the availability of this production trace that is thought to guide test performance. Pictures are thought to be distinctive because they contain many unique visual features and it is the availability of these features that guides test performance. Even though the encoding processes through which the distinctive elements are incorporated into the relevant memory trace may be different for the production effect and the picture superiority effect, the availability of these distinct elements makes the retrieval process capable of probabilistically differentiating studied items from distractors items to benefit memory performance. As described above, if the participant were to remember producing a word or were to recognise the visual elements of a picture, the perceived likelihood of that item being a distractor is low: It is unlikely that participants would recently have produced a distractor item or that they would recently have encountered a picture containing the same unique visual features. It thus seems reasonable to hypothesise that the combination of distinctive features across multiple representational dimensions would be especially effective at differentiating studied items from distractors. This is because the probability of having recently produced *and* visually inspected a similar distractor item is especially unlikely. The current

[1] Dodson and Schacter (2001) referred to the use of a production trace to infer whether an item had been studied as the "distinctiveness heuristic".

[2] We would like to thank Dr Jason Ozubko for suggesting this framing.

manuscript explores this possibility. We predicted that producing the name of a picture would result in an especially distinctive memory trace such that there would be a significantly larger production effect for pictures than for words.

Experiment 1 presented participants with two blocked lists of pictures and words, a random half of which participants were instructed to mouth quietly (produced condition) and the other half of which they were instructed to read silently without concurrent motion or utterance (non-produced condition). Following the presentation of all study items, participants completed a speeded yes-no recognition task to discriminate the ''old'' study items from ''new'' foil items. Experiment 2 replicated the methods of Experiment 1, but used an intermixed presentation of pictures and words at both study and test rather than a blocked presentation. Experiment 3 replicated Experiment 2, but used a self-paced recognition task rather than a speeded recognition task. To foreshadow our results, across all three experiments there was a significant production and picture superiority effect in addition to the predicted two-way interaction.

EXPERIMENT 1

In Experiment 1 participants were presented with the blocked presentation of pictures and words. Colour was used to instruct participants to either silently name the picture/word or to explicitly produce the picture/word by mouthing it. Pictures were surrounded by a red or blue border and words were presented with a red or blue font. Pictures and words were presented in separate blocks at study and test to minimise any tendency for participants to adopt a visualisation or self-generated imagery strategy as a means of encoding the words. Immediately following the study phase we administered a speeded yes-no recognition task to evaluate memory for each studied item as well as the false alarm rates for an equal number of foil items (i.e., items not presented at study). We adopted a speeded (as opposed to self-paced) procedure as a pre-emptive measure to ensure that performance did not approach ceiling in the picture-mouthed condition; this does not obviate the fact that sensitivity and response bias were our dependent measures of interest.

Method

Participants. A total of 32 undergraduate students (13 male, 19 female) participated in exchange for course credit. The experiment was run in a single session lasting approximately 30 minutes. All participants reported normal or corrected-to-normal vision and a good understanding of the English language.

Stimuli and apparatus. All experimental procedures were presented using custom software developed in the Python programming language (www.python.org) with the Pygame development library (www.pygame.org) loaded on either a 15-inch 1440 × 900-resolution MacBook Pro running Mac OS X 10.6 or a 24-inch iMac computer running Mac OSX Leopard, version 10.5. Responses were recorded via a built-in laptop keyboard (in the case of the MacBook Pro) or a standard Macintosh Universal Serial Bus keyboard (in the case of the iMac). Words and fixation stimuli were printed in Arial size 42 font.

The stimuli consisted of 200 two-dimensional coloured line drawings and their corresponding names (e.g., barn, frog) sampled from Rossion and Pourtois (2004) who reported high correspondence between the picture and word stimuli. These pictures are available for download from http://www.nefy.ucl.ac.be/facecatlab/stimuli.htm. All pictures were 16-bit RGB colour, had a resolution of 72 × 72 dpi, and were converted to JPEG format. At both study and test, pictures and words were presented at centre and were either surrounded by a red (RGB: 255,0,0) or blue (RGB: 0,0,255) border (pictures) or printed with a red or blue font (words). The border/ink colour served as the production instruction. For each participant the stimuli were randomly distributed across each combination of production (mouthed, silent), stimulus (picture, word), and item type (study, foil) resulting in eight lists, each containing 25 items.

Procedure. Participants were told that they would be presented with a series of pictures and words, each of which would be surrounded by a red/blue border or printed with a red/blue font, respectively. Half of the participants were instructed to silently mouth the name of each ''blue'' study item and to name each ''red'' study item in their head without any associated motion or utterance; these instructions were reversed for the remaining participants.[3] The distinction between silently mouthing and mentally naming an item was demonstrated for the participants to ensure they understood the instructions. Participants were instructed that both pictures and words would be presented in separate blocks, and that the order in which these blocks occurred would be the same for both study and test. For half of the participants the picture block was presented first and for the other half of the participants the word block was presented first. The blue and red borders/fonts (i.e., the production instructions) were randomly interspersed within each of these blocks, with the restriction that half of the items in each block were presented with blue and half were presented with red.

Study phase. Each study phase trial consisted of a fixation stimulus ("+") lasting 500 ms followed by the study item for 2000 ms. On any given trial the study item consisted of either a picture or a word surrounded by or printed in the colour red or blue (depending on the production condition). Based on the associated border/font colour, participants responded by naming the study item in their head or by silently mouthing the name of the study item. There were a total of 100 study phase trials representing 25 replications of each of the following conditions: picture-mouthed, picture-silent, word-mouthed, and word-silent.

Recognition phase. Following the study phase participants were tested for their memory of the preceding study items using yes-no recognition. Prior to beginning this phase the researcher provided further instructions. Participants were informed that they would be presented with all of the pictures and words from the study phase as well as an equal number of "new" pictures and words that they had not studied (i.e., foils). Each test item was presented one at a time in separate picture and word blocks. The blocks were presented in the same order as at study, although the individual items within each block were randomised. Participants were asked to indicate whether each item was an "old" studied item or a "new" foil item. Half of the participants were instructed to press the "f" key to indicate that "yes" they recognised the corresponding test item and to press the "j" key to indicate that "no" they did *not* recognise the corresponding test item; this response assignment was reversed for the remaining participants. Participants were instructed that they would have limited time to respond to each test item and should therefore respond as accurately but also as quickly as possible.

Because we intended to conduct a signal detection analysis to investigate how the production and picture superiority effects influence both sensitivity and response criterion, it was important that we estimate separate foil false alarm rates for each condition (see Macmillan & Creelman, 2005). To ensure separate false alarm rates for each condition we maintained the colour coding from the study phase such that items studied in blue were tested in blue and that items studied in red were tested in red. We also equally distributed the colour assignment to the foil pictures and words.

Each recognition phase trial began with a fixation stimulus ("+") for 500 ms after which the test item was presented and remained onscreen for 2000 ms. Participants were required to make a yes/no response during this period. Immediately following the presentation of each test item, the next trial began. No feedback was given regarding the speed or accuracy of the response. There were a total of 200 trials within the recognition phase representing 25 replications of each of the following eight conditions: picture-blue-study, picture-red-study, word-blue-study, word-red-study, picture-blue-foil, picture-red-foil, word-blue-foil, and word-red-foil (the naming convention is such that the first word refers to the stimulus, the second to the border/font colour, and the last to whether the item was presented as a study or foil item).

Results

Although the speed of making the recognition decision was not of primary interest, we have nevertheless depicted the reaction times (RTs) in Table 1 for the interested reader. Note that different numbers of trials contribute to these means due to empty cells for some participants (e.g., some participants did not produce false alarms to foils for all conditions). These data were not analysed.

[3] Mouthing was selected as our production manipulation over "speaking aloud" because participants were run in adjacent rooms without any soundproofing. We were concerned that other participants would hear any vocalisation of the study items. MacLeod et al. (2010) have found mouthing to result in a reliable production effect so we do not expect this methodological feature to impact our interpretation.

TABLE 1
Mean reaction times

	Stimulus type			
	Pictures		*Words*	
	Hits	*False alarms*	*Hits*	*False alarms*
Experiment 1				
Mouthed	649 (2)	726 (4)	662 (4)	696 (5)
Silent	681 (3)	772 (9)	610 (5)	707 (7)
Experiment 2				
Mouthed	699 (3)	750 (5)	712 (3)	744 (4)
Silent	757 (3)	763 (6)	760 (4)	795 (5)
Experiment 3				
Mouthed	1055 (10)	1520 (33)	1080 (9)	1191 (11)
Silent	1195 (10)	1524 (23)	1224 (10)	1246 (14)

Mean reaction times (ms) for hits and false alarms for Experiments 1–3 as a function of Production (Mouthed, Silent) and Stimulus Type (Pictures, Words); standard error is provided in parentheses.

The mean hits and false alarms are shown in Table 2. Using the procedure described by Macmillan and Creelman (2005), sensitivity (d') and response bias (C) were calculated on a participant-by-participant basis and analysed as a function of production (mouthed, silent) and stimulus (picture, word) using separate repeated-measures ANOVAs. A d' of 0 represents chance performance; positive values represent greater separation between the signal and noise distributions favouring the signal; negative values represent greater separation between the signal and noise distributions favouring the noise. Values of C range from liberal (requiring less "signal" to classify a test item as "old"; $C < 0$) to conservative (requiring more "signal" to classify a test item as "old"; $C > 0$). The sensitivity and response bias data are depicted in Figure 1. An analysis of the raw performance data or alternative measures of sensitivity and bias (e.g., A' and B''_D) demonstrated a pattern comparable to that described below for d' and C. Although the analyses reported below initially included block as a factor (i.e., pictures first, words first), this factor was removed for the sake of exposition because it did not result in a main effect and did not interact with stimulus and/or production (all ps $> .40$).

TABLE 2
Mean hits and false alarms

	Stimulus type			
	Pictures		*Words*	
	Hits	*False alarms*	*Hits*	*False alarms*
Experiment 1				
Mouthed	.84 (.02)	.11 (.02)	.64 (.03)	.31 (.03)
Silent	.65 (.03)	.09 (.01)	.44 (.03)	.19 (.03)
Experiment 2				
Mouthed	.83 (.02)	.13 (.02)	.64 (.03)	.25 (.03)
Silent	.57 (.03)	.11 (.02)	.40 (.04)	.17 (.02)
Experiment 3				
Mouthed	.83 (.02)	.09 (.01)	.64 (.03)	.21 (.03)
Silent	.68 (.03)	.11 (.02)	.47 (.03)	.14 (.02)

Mean hits and false alarms for Experiments 1–3 as a function of Production (Mouthed, Silent) and Stimulus Type (Pictures, Words); standard error is provided in parentheses.

Sensitivity. The main effect of production was significant, $F(1, 31) = 15.09$, $MSe = 0.25$, $p < .01$, $\eta^2_g = .06$, indicating greater sensitivity to items in the mouthed condition ($M = 1.62$, $SE = 0.10$) than to items in the silent condition ($M = 1.29$, $SE = 0.11$). The main effect of stimulus was also significant, $F(1, 31) = 145.67$, $MSe = 0.29$, $p < .01$, $\eta^2_g = .41$, indicating greater sensitivity to pictures ($M = 2.03$, $SE = 0.11$) than to words ($M = 0.88$, $SE = 0.10$). The critical two-way interaction between production and stimulus was also significant, $F(1, 31) = 4.60$, $MSe = 0.29$, $p < .04$, $\eta^2_g = .02$; as depicted in Figure 1, the magnitude of the production effect was larger for pictures compared to words.

Response bias. The main effect of production was significant, $F(1, 31) = 61.29$, $MSe = 0.10$, $p < .01$, $\eta^2_g = .27$, indicating a more liberal response bias when responding to items in the mouthed condition ($M = 0.12$, $SE = 0.05$) relative to items in the silent condition ($M = 0.56$, $SE = 0.05$). Neither the main effect of stimulus, $F(1, 31) = 0.87$, $MSe = 0.13$, $p > .35$, $\eta^2_g < .01$, nor the production × stimulus interaction, $F(1, 31) = 2.02$, $MSe = 0.07$, $p > .16$, $\eta^2_g < .01$, approached significance.

Discussion

The purpose of Experiment 1 was to determine whether the production effect and picture superiority effect would interact. We obtained a production effect and a picture superiority effect, as well as the predicted two-way interaction of these two effects. Our findings are consistent with our

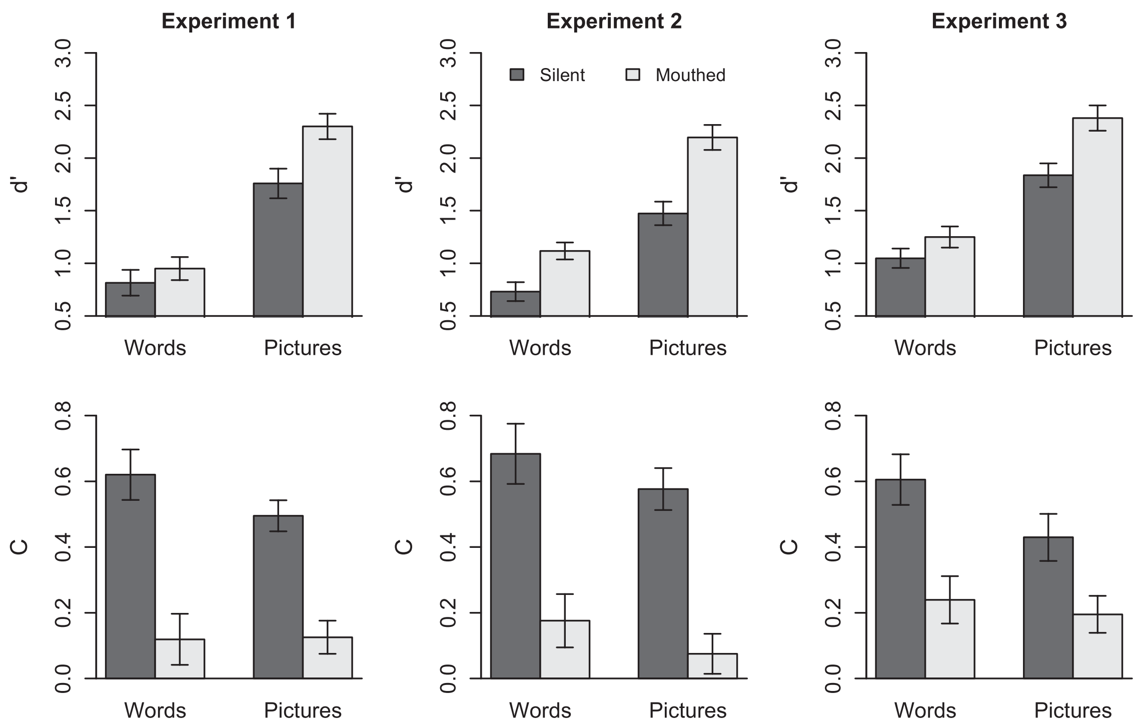

Figure 1. Sensitivity (d') and Response Bias (C) for Experiments 1–3 as a function of Production (Mouthed, Silent) and Stimulus Type (Pictures, Words); error bars represent one standard error of the mean.

expectation that access to both a production trace and unique visual features is especially advantageous for subsequent memory performance.

Interestingly we observed the unexpected finding that the production effect (but not the picture superiority effect) influenced the response criterion used to discriminate ''old'' from ''new'' items within the recognition phase. Participants responded more liberally to items in the mouthed condition than to items in the silent condition (see also Dodson & Schacter, 2001). This finding is important to consider when conducting future studies of the production effect using recognition memory. Traditional analyses of recognition performance often focus on hits while de-emphasising false alarms and thereby conflate sensitivity and response bias. Given that the current results suggest that participants respond more liberally to mouthed items, this would act to inflate the magnitude of the production effect in an analysis of hit rates. This point is depicted visually in Figure 2, which represents the relation between response criterion and the magnitude of the production effect using hits while assuming different hypothetical effects for sensitivity. When the produced and non-produced conditions demonstrate a similar response criterion, differences in the hits for produced and non-produced items are related to differences in the sensitivity for produced and non-produced items. As the response criterion in the produced condition becomes more liberal, the production effect is magnified in the hit rates: When the difference

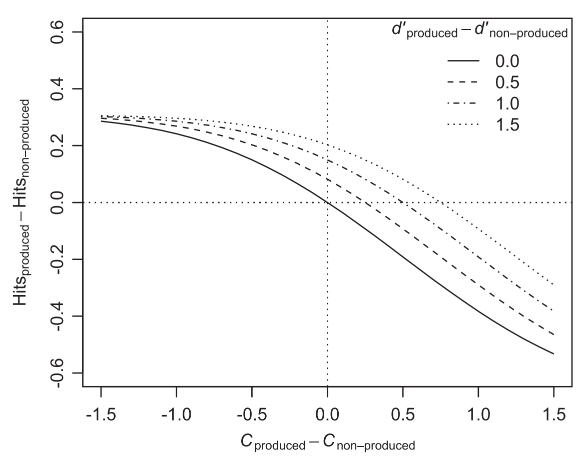

Figure 2. The mathematical relationship between differences measured by ''Hits'' and hypothetical changes in response bias (C) across assumed differences in sensitivity (d'). The values of $d'_{non\text{-}produced}$ and $C_{non\text{-}produced}$ have been set as 1 and 0; however, the same relationship is true when these values are varied.

in d' is 0.50 and the difference in response criterion is 0, the magnitude of the production effect measured by hits is .08, whereas when the difference in d' is 0.50 and the difference in response criterion is -0.50 (i.e., more liberal), the magnitude of the production effect measured by hits increases to .20. Conversely, as the produced condition becomes more conservative, the production effect is masked in the hit rates: When the difference in d' is 0.50 and the difference in response criterion is 0.50 (i.e., more conservative), the magnitude of the production effect measured by hits becomes $-.09$ (a reverse production effect).

Although not an a priori prediction, the liberal response bias observed for produced compared to non-produced items may be interpreted in relation to the distinctiveness account. MacLeod et al. (2010) argued that producing an item results in the accrual of unique motor and/or sensory information that may be reconstituted at test to guide performance. Because even a partial production trace may be enough to influence recognition decisions, participants in our study might have interpreted access to *any* production of a particular picture or word—including those productions that occurred outside the context of the experiment—as evidence of having studied the item. Insofar as this approach was applied to the relevant distractor items, a liberal response bias would be expected. For example, a given participant might be able to recall producing "dog" but misattribute that production trace to the study phase, when in fact it occurred *preceding* the experiment. Participants would then need to discriminate the source of the production trace, potentially biasing them towards false alarms in these cases. However, such cases should only occur for distractors printed in or surrounded by the colour associated with production at study: Participants were aware that the colours at test corresponded to those at study (i.e., colour assignment did not change between phases) and therefore they could dismiss any production trace associated with distractors outlined or printed in the colour that signalled non-produced items. The end result would be more "yes" responses to (study or foil) items within the condition associated with mouthing compared to the condition associated with silent reading, and therefore the relatively more liberal response bias that we observed in this experiment.

EXPERIMENT 2

In Experiment 1 we obtained a production effect and a picture superiority effect as well as the predicted interaction. Experiment 2 was intended to replicate this finding while intermixing the pictures and words to increase the relative distinctiveness of the items, and to generalise our findings to previous work (e.g., MacLeod et al., 2010). The risk is that intermixing pictures and words will encourage a mental imaging strategy for words that will functionally equate the processing of pictures and words, and thereby eliminate the picture superiority effect. However, based on the results of Experiment 1 we think that this is unlikely. If the presentation of pictures had prompted the use of an imaging strategy when presented with words, there would have been an effect of block order in Experiment 1. Given that there was not, we were motivated to repeat Experiment 1 using an intermixed presentation of pictures and words. Thus, in the study phase of Experiment 2, pictures and words outlined or printed in blue or red were intermixed and presented randomly in a single block of trials. Immediately following the study phase, we administered a speeded yes-no recognition task to evaluate memory for each studied item as well as the false alarm rates for an equal number of foil items.

Method

Participants. A total of 36 undergraduate students (13 male, 23 female) participated in exchange for course credit. The experiment was run in one session lasting approximately 30 minutes. All participants reported normal or corrected-to-normal vision and a good understanding of the English language.

Stimuli and apparatus. The stimuli and apparatus used in Experiment 2 were identical to those used in Experiment 1.

Procedure. The procedure for Experiment 2 was identical to that of Experiment 1 with the exception that the pictures and words were randomly intermixed (as opposed to blocked) throughout both the study and test phases.

Results

As was the case in Experiment 1, the speed to make the recognition decision was not of primary interest. Nevertheless, recognition RTs are depicted in Table 1, again with the caveat that different numbers of trials contribute to the means due to empty cells. These data were not analysed.

Mean hits and false alarms for the recognition test are provided in Table 2. Sensitivity (d') and response bias (C) were calculated and analysed as a function of production (mouthed, silent) and stimulus (picture, word) using separate repeated-measures ANOVAs.

Sensitivity. As in Experiment 1 the main effect of production was significant, $F(1, 35) = 74.21$, $MSe = 0.15$, $p < .01$, $\eta^2_g = .36$, indicating greater sensitivity to items in the mouthed condition ($M = 1.66$, $SE = 0.09$) than to items in the silent condition ($M = 1.10$, $SE = 0.08$). The main effect of stimulus was also significant, $F(1, 35) = 126.11$, $MSe = 0.24$, $p < .01$, $\eta^2_g = .37$, indicating greater sensitivity to pictures ($M = 1.83$, $SE = 0.10$) than to words ($M = 0.92$, $SE = 0.07$). As depicted in Figure 1, the interaction was significant, $F(1, 35) = 5.21$, $MSe = 0.19$, $p < .03$, $\eta^2_g = .02$, reflecting a larger production effect for pictures compared to words.

Response bias. As in Experiment 1 the main effect of production was significant, $F(1, 35) = 71.92$, $MSe = 0.13$, $p < .01$, $\eta^2_g = .24$, indicating a more liberal response bias when responding to items in the mouthed condition ($M = 0.13$, $SE = 0.06$) relative to items in the silent condition ($M = 0.63$, $SE = 0.07$). Neither the main effect of stimulus, $F(1, 35) = 3.18$, $MSe = 0.12$, $p > .07$, $\eta^2_g = .01$, nor the production $\times$ stimulus interaction, $F(1, 35) = 0.01$, $MSe = 0.05$, $p > .93$, $\eta^2_g < .01$, reached significance.

Discussion

In Experiment 2 we observed a complete replication of the findings of Experiment 1: Both the picture superiority effect and production effect were evident in the analysis of the d' data but only the production effect influenced response bias. The production $\times$ stimulus interaction was again significant indicating a larger production effect for pictures compared to words.

EXPERIMENT 3

Both Experiments 1 and 2 produced the predicted production $\times$ stimulus interaction. Experiment 3 replicated Experiment 2 with the exception that the speeded recognition task was replaced with a self-paced recognition task. The purpose of this change was to ensure that the measured effects and interactions were not influenced by the demands associated with the speeded response at recognition. Further, we had expected that permitting additional time to consider the distinctive features of the test items could potentially mitigate the effect of production on response bias. It is possible that the speeded recognition task used in Experiments 1 and 2 might have rushed source analysis of the production trace associated with any given test item. If participants recalled recently producing a test item outside the experiment, an especially rapid response might not have provided the time necessary to resolve the source of the production trace, thereby accounting for the more liberal response bias overall.

As in Experiment 2 pictures and words outlined or printed in blue or red were intermixed and presented randomly during the study phase. Immediately following the study phase we administered a self-paced yes-no recognition task to evaluate memory for each studied item as well as the false alarm rates for an equal number of foil items.

Method

Participants. A total of 36 undergraduate students (9 male, 27 female) participated in exchange for course credit. The experiment was run in a single session lasting approximately 30 minutes. All participants reported normal or corrected-to-normal vision and a good understanding of the English language.

Stimuli and apparatus. The stimuli and apparatus used in Experiment 3 were identical to those used in Experiments 1 and 2.

Procedure. The procedure for Experiment 3 was identical to that of Experiment 2 with the exception that during the recognition phase each test item remained on-screen until a response was made.

Results

Even though the recognition task in Experiment 3 was not speeded, we nevertheless recorded the response times; these are shown in Table 1. As was the case for Experiments 1 and 2, these data are based on different numbers of trials due to empty cells and were not analysed.

The mean hits and false alarms for the recognition test are shown in Table 2. Sensitivity (d') and response bias (C) were once again calculated and analysed as a function of production (mouthed, silent) and stimulus (picture, word) using separate repeated-measures ANOVAs. Sensitivity and response bias are depicted in Figure 1.

Sensitivity. The main effect of production was significant, $F(1, 35) = 16.81$, $MSe = 0.30$, $p < .01$, $\eta_g^2 = .08$, indicating greater sensitivity to items in the mouthed condition ($M = 1.81$, $SE = 0.10$) than to items in the silent condition ($M = 1.44$, $SE = 0.09$). The main effect of stimulus was also significant, $F(1, 35) = 127.59$, $MSe = 0.26$, $p < .01$, $\eta_g^2 = .36$, indicating greater sensitivity to pictures ($M = 2.11$, $SE = 0.10$) than to words ($M = 1.15$, $SE = 0.08$). As in both Experiments 1 and 2 the predicted interaction between production and stimulus was also significant, $F(1, 35) = 5.54$, $MSe = 0.20$, $p < .03$, $\eta_g^2 = .02$, with a larger production effect for pictures than words.

Response bias. Only the main effect of production was significant, $F(1, 35) = 28.02$, $MSe = 0.12$, $p < .01$, $\eta_g^2 = .12$, indicating a more liberal response bias when participants responded to items in the mouthed condition ($M = 0.22$, $SE = 0.05$) than when they responded to items in the silent condition ($M = 0.52$, $SE = 0.06$). Neither the main effect of stimulus, $F(1, 35) = 3.02$, $MSe = 0.14$, $p > .09$, $\eta_g^2 = .02$, nor the production × stimulus interaction, $F(1, 35) = 2.27$, $MSe = 0.07$, $p > .14$, $\eta_g^2 < .01$, was significant.

Discussion

In addition to both a production effect and a picture superiority effect, the results of Experiment 3 again revealed the predicted interaction of production and stimulus: The production effect was significantly larger for pictures than for words. Experiment 3 replicated the methods of Experiment 2 using a self-paced as opposed to a speeded yes-no recognition task to allow more time for the participant to evaluate the distinctive features associated with the test items. We hypothesised that requiring a speeded response would leave little time for the participant to perform a fine-grained source analysis of any active production trace at test. If the effect of production on response bias were thus caused by the misattribution of pre-experimental production traces, we would expect this difference to be of larger magnitude in the context of a speeded recognition task than a self-paced recognition task. To determine whether this was the case we compared the response bias difference for the mouthed versus silent conditions in Experiments 1 and 2 (which used speeded tasks)[4] to that obtained in Experiment 3 (which used a self-paced task). This analysis revealed that the magnitude of the difference in response bias for the mouthed condition versus the silent condition was larger for the speeded recognition task (0.47) than for the self-paced recognition task, 0.30; $t(102) = 2.46$, $p < .02$. This difference in magnitude is almost entirely attributable to a shift towards more liberal responding within the mouthed condition, as opposed to more conservative responding within the silent condition (see Figure 1).

GENERAL DISCUSSION

The purpose of the current study was to determine whether the production effect and the picture superiority effect would interact, such that the production effect would be larger for pictures than for words. Both effects have been attributed to distinctive processes applied at encoding and used to guide memory performance at test (see Dodson & Schacter, 2001; Fawcett, 2012; Lloyd & Miller, 2011). Although these processes are similar they act on different dimensions, and when combined may be especially effective at differentiating studied items from

[4] Comparison of the magnitude of the difference in response bias for the mouthed condition versus the silent condition in Experiments 1 and 2 revealed no difference, $t(66) = 0.84$, $p > .40$.

distractors. Therefore we expected the two effects to interact such that the production of a picture would result in an especially distinctive memory trace and, as a result, superior recognition. Analysis of the findings reported in Figure 1 support this claim: The production effect was significantly larger for pictures compared to words across all three experiments.

Future researchers must remain mindful of how production influences both sensitivity *and* response bias when interpreting recognition memory performance. This is depicted in Figure 2, which represents the relation between response bias and the magnitude of the production effect using hits as opposed to sensitivity. As already discussed, this figure demonstrates that changes in response bias can magnify (liberal response bias) or even reverse differences (conservative response bias) measured using hits alone, even though sensitivity remains unchanged. For this reason hits should not be used as a dependent measure unless there is good reason to believe that the independent variable does not affect response bias in addition to sensitivity (as noted earlier, when response bias is comparable across conditions differences in hits are related to differences in sensitivity).

While sensitivity should be calculated directly whenever possible, measures of sensitivity require separate false alarm rates for each experimental condition, and thus when only a single false alarm rate is available the analysis of sensitivity and response bias is confounded with the analysis of hits and false alarms. This is a concern because most production experiments to date have measured only a single false alarm rate (e.g., MacLeod et al. , 2010) with the exception of research manipulating production between-participants as opposed to within-participants. Of these between-participants studies, only Dodson and Schacter (2001) report response bias (in this case B''_D). Consistent with our findings they observed a trend (not significant) favouring a more liberal bias in the produced condition. As an attempt to control for any contributions associated with hearing the study item as it was produced, Dodson and Schacter (2001) used auditory study items for the non-produced condition, and therefore their findings should be considered with caution when extrapolating to the more typical comparisons with a silent control condition. Although there is little evidence as to whether production affects response bias in a typical within-participants paradigm that uses a read-silently control condition, based on the results of our three experiments we would certainly expect this to be the case. The fact that most studies are unable to address this concern because they measure only a single false alarm rate is problematic. Figures 1 and 2 clearly depict the need to separate foil false alarm rates for produced and non-produced items, to ensure that the production effect is measured accurately.

Beth, Budson, Waring, and Ally (2009) have presented evidence that participants also respond more liberally to pictures compared to words. Although the effect of stimulus on response bias failed to reach significance in any of the current experiments, it approached significance in both Experiments 2 and 3. Indeed, a visual inspection of Figure 1 reveals a consistent trend towards a more liberal response bias for pictures compared to words across all three experiments of the current study. Analysis of the combined response bias data from Experiments 1, 2, and 3 reveals this effect to be significant, $F(1, 103) = 6.82$, $MSe = 0.06$, $p < .02$, $\eta^2_g = .02$, albeit very small in relation to the effect of production on response bias ($\eta^2_{g\text{-Production}} = .21$, .24 and .12 in Experiments 1, 2, and 3). Experiment does not itself interact with the effect (nor does production if included) and limiting this analysis to Experiments 2 and 3 where the effect was marginally significant still results in a significant effect of comparable magnitude ($\eta^2_g = .02$). So, although participants may respond more liberally to pictures ($M = 0.32$, $SE = 0.03$) compared to words ($M = 0.41$, $SE = 0.04$), the fact that this effect is relatively small suggests that using hit rates to measure the picture superiority effect likely suffers from relatively less contamination from response bias than using hit rates to measure the production effect (see Figure 2). Nonetheless, studies comparing memory for pictures and words naturally produce separate false alarm rates allowing the calculation of d' and C.

Throughout this study we have adopted the theoretical position that distinctiveness underlies both the production effect and the picture superiority effect. We have described each effect as arising from the application of distinctive encoding processes at study (production or evaluation of unique visual features) resulting in the availability of distinct retrieval cues to guide performance at test. However, we would be remiss to not mention that alternative accounts exist for both the production effect as well as the picture superiority effect. Most relevant to the current investigation is the notion that producing a study item or viewing it as a picture (as opposed to a word) results in a stronger memory trace. Such a single-process account makes no assumption as to the retrieval strategies employed at test but rather focuses on the general benefits of additional, deeper processing applied at study (see also Craik & Lockhart, 1972).

To the extent that the strength of a memory trace is tied to its elaboration within memory and therefore to the probability of possessing one or more features capable of guiding recognition, there is perhaps less tension between the single-process model and the other distinctiveness accounts than might first appear. Even so, the weight of current evidence appears to support a specific role for distinctiveness within the production effect. First, the production effect has been found to be less robust between-participants as compared to within-participants (for discussion, see Fawcett, 2012; MacLeod et al., 2010). If production merely enhanced the strength of the memory trace, study design should not matter. Ozubko and MacLeod (2010) also found that the production effect disappears when participants are instructed to read aloud the foil items prior to test. Doing so reduces the discriminative value of any given production trace because the presence of such a trace at test could signal a produced study item *or* a produced foil item. Once again, a purely strength-based account would not predict such an outcome. Finally, a purely strength-based account would not predict the production $\times$ stimulus interaction observed for sensitivity in the current experiments or the impact of recognition task (speeded vs self-paced) on response bias. Instead these findings are most compatible with the view that participants are using distinct features to discriminate between the ''old'' and ''new'' test items.

In conclusion, the current experiments demonstrate that the production effect is larger for pictures than for words. We have argued that this interaction is most readily explained with reference to distinctiveness at encoding that can be used to guide subsequent recognition performance at test. The interaction of the production and picture superiority effects suggest that the application of multiple distinctive processes produces benefits greater than the sum of the individual processes. As an unexpected finding we have also determined that participants respond more liberally to produced than to non-produced items. This finding could have important implications for the use of hit rates to measure the production effect (i.e., because they may over- or under-estimate the magnitude of the actual effect). While there was also a tendency for participants to respond more liberally to pictures than to words, the fact that this effect was small and became significant only in our pooled data suggests that the implications are less severe for the use of hit rates to measure the picture superiority effect.

Manuscript received 28 November 2011
Manuscript accepted 7 May 2012
First publshed online 27 June 2012

REFERENCES

Beth, E. H., Budson, A. E., Waring, J. D., & Ally, B. A. (2009). Response bias for picture recognition in patients with Alzheimer's disease. *Cognitive and Behavioural Neurology*, *22*(4), 229–235.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684.

Dodson, C. S., & Schacter, D. L. (2001). ''If I had said it I would have remembered it'': Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review*, *8*, 155–161.

Fawcett, J. M. (2012). *The production effect benefits performance in between-subject designs: A meta-analysis*. Manuscript submitted for publication.

Hunt, R. R. (2006). The concept of distinctiveness in memory research. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 3–26). New York, NY: Oxford University Press.

Hourihan, K. L., & Macleod, C. M. (2008). Directed forgetting meets the production effect: Distinctive processing is resistant to intentional forgetting. *Canadian Journal of Experimental Psychology*, *62*(4), 242–246.

Lloyd, M. E., & Miller, J. K. (2011). Are two heuristics better than one? The fluency and distinctiveness heuristics in recognition memory. *Memory & Cognition*, *39*, 1264–1274. DOI: 10.3758/s13421-011-0093-0.

MacLeod, C. M. (2010, June). *When learning met memory*. Paper presented at the annual meeting of the Canadian Society for Brain, Behavior and Cognitive Science, Halifax, Nova Scotia.

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 671–685.

Macmillan, N., & Creelman, C. D. (2005). *Detection theory: A user's guide*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Madigan, S. (1983). Picture memory. In J. C. Yuille (Ed.), *Imagery, memory, and cognition: Essays in honor of Allan Paivio* (pp. 65–89). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Ozubko, J. D., & MacLeod, C. M. (2010). The production effect in memory: Evidence that distinctiveness underlies the benefit. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *36*, 1543–1547.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Journal of Canadian Psychology*, *45*, 255–287.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, *33*, 217–236.

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language*, *40*, 1–24.