Psychonomic Bulletin & Review
https://doi.org/10.3758/s13423-024-02566-5

BRIEF REPORT

Generation (not production) improves the fidelity of visual representations in picture naming

Jedidiah W. Whitridge[1] · Chris A. Clark[1] · Kathleen L. Hourihan[1] · Jonathan M. Fawcett[1,2]

Accepted: 18 June 2024

Abstract

The production effect refers to the finding that participants better remember items read aloud than items read silently. This pattern has been attributed to aloud items being relatively more distinctive in memory than silent items, owing to the integration of additional sensorimotor features within the encoding episode that are thought to facilitate performance at test. Other theorists have instead argued that producing an item encourages additional forms of processing not limited to production itself. We tested this hypothesis using a modified production task where participants named monochromatic line drawings aloud or silently either by generating the names themselves (no label condition) or reading a provided label (label condition). During a later test, participants were presented with each line drawing a second time and required to reproduce the original color and location using a continuous slider. Production was found to improve memory for visual features, but only when participants were required to generate the label themselves. Our findings support the notion that picture naming improves memory for visual features; however, this benefit appears to be driven by factors related to response generation rather than production itself.

Keywords Memory · Production · Distinctiveness · Generation · Processing

Introduction

Scientists and philosophers alike have long sought to catalogue the strategies and mechanisms through which studied information is best secured in memory (e.g., Bellezza, 1981; Hebb, 1949; James, 1890). One rather old strategy (Ekstrand et al., 1966; for a modern review, see MacLeod & Bodner, 2017) that has attained renewed notoriety in recent years involves the overt production of certain study items – for example by reading them aloud (e.g., Hopkins & Edwards, 1972; MacLeod et al., 2010), singing them (e.g., Quinlan & Taylor, 2013, 2019), or mouthing them (e.g., Forrin et al., 2012). Each form of production involves additional preparatory or sensorimotor stages that some allege prove beneficial at later test (Fawcett et al., 2012; Forrin et al., 2012). This strategy has been shown to be effective, with produced items better recalled (e.g., Lin & MacLeod, 2012) or recognized (e.g., Conway & Gathercole, 1987) than other, non-produced items (e.g., read silently); this has been termed the *production effect* (MacLeod et al., 2010).

Modern theoretical explanations of the production effect have largely centred on the notion of distinctiveness (e.g., MacLeod et al., 2010, 2022; Ozubko & MacLeod, 2010; Ozubko et al., 2014). This framework contends that production (e.g., reading the word aloud) evokes additional processes and representational elements (e.g., motor preparation, audition) not otherwise present during non-productive study; these processes become bound to the encoding episode, representing what some refer to as the production trace (Fawcett, 2013). In this manner, produced items "stand out" against the backdrop of non-produced items in memory, making them more readily accessible at retrieval, allowing the participant to use access to the production trace to discriminate between studied and unstudied items (MacLeod et al., 2010). Early support for this account came from the apparent finding that production only influenced memory when manipulated within rather than between subjects (e.g., Hopkins & Edwards, 1972; MacLeod et al., 2010), and also that the effect could be eliminated by manipulations that diminished the utility of the production trace as a means of discriminating between old and new items (Ozubko & MacLeod, 2010).

✉ Jonathan M. Fawcett
jfawcett@mun.ca

[1] Department of Psychology, Memorial University of Newfoundland, St John's, NF, Canada

[2] Psychology Department, Memorial University of Newfoundland, St John's, NL, Canada

However, the presence of a between-subject production effect in recognition memory (for a review, see Fawcett et al., 2023)[1] challenges the notion that this effect solely arises from distinctiveness, given that there are no silent items to provide the "backdrop" against which the aloud items might "stand out." Further characterizing the underlying processes, Fawcett and Ozubko (2016) demonstrated that the between-subject production effect represents improved familiarity alone, whereas the within-subject production effect represents improved familiarity and recollection. They speculated that strategic use of the production trace might account for the recollective component whereas variation in attention or engagement might account for the familiarity component (see also Ozubko et al., 2012). The idea that participants are more engaged during aloud trials has been supported by applied work showing a reduction in mind wandering while reading passages aloud versus silently (Varao Sousa et al., 2013), electrophysiological evidence showing modulation of early indices of attentional engagement even preceding the productive act (e.g., the P300; Hassall et al., 2016; Zhang et al., 2023), neuroimaging evidence demonstrating increased activation in brain regions consistent with attentional regulation and semantic processing during production (Bailey et al., 2021), behavioral evidence that fluctuating – but not continuous – background noise eliminates the benefit of production (Mama et al., 2018), and pupillometric evidence revealing changes in mental effort predictive of the magnitude of the production effect even preceding word onset (Willoughby et al., 2019). Together, these findings suggest that participants process to-be-produced items in a fundamentally different manner than those intended for non-productive study, characterized by heightened attentional engagement and sensitivity to the encoded material and its context.

Past research provides evidence for production-related memory improvements extending beyond the boundaries of the chosen productive modality. For example, the production effect persists despite – and is even unaffected by – efforts to eliminate the utility of the production trace by requiring a two-alternative forced choice with matched homophones (e.g., WHALE-WAIL) at test that would obviate the value of the production trace (Fawcett et al., 2022). Several studies have shown a production effect for nameable objects (e.g., Fawcett et al., 2012; MacLeod et al., 2022; Zormpa et al., 2019a, b), including tasks putatively dependent on visual details (e.g., Hourihan & Churchill, 2020; Richler et al., 2013); however, none have thus far investigated how production influences the probability of recalling visual details or the precision with which those details are represented. The present study was designed to address this question.

Current experiments

Two experiments were conducted to evaluate whether production improves the probability and/or fidelity of visual representations; in either, participants completed a modified production task involving the naming of colored objects with a continuous color and location judgment at test. These experiments differed only with respect to whether a verbal label was provided at study. For the sake of brevity, we report and model these data together.[2] The combined data set was analyzed using two parametric models estimating either the fidelity (i.e., precision) of the respective judgment (von Mises model) or quantifying separately the probability of recollecting the color or location and the fidelity with which those features were represented when accessible (mixture model; e.g., Fawcett et al., 2016).

A strict interpretation of the typical distinctiveness account would predict production to have no impact on memory for visual detail because the production trace is itself orthogonal to such information and memory of having named the object aloud would have no diagnostic value (as all test items will have been studied). If differences emerged, they would instead imply variation in how the objects named aloud or silently were encoded during the initial phase. More specifically, a production effect on visual detail would suggest that producing items alters the manner in which participants engage with or attend to those items, congruent with attentional framework of the effect. However, because picture naming – in the absence of a verbal label – requires generation as a precursor to production (Zormpa et al., 2019a), the emergence of a production effect in the no-label condition would conflate the benefits of production with those of

[1] Although between-subjects production benefits for overall recall appear unreliable (e.g., Jones & Pyc, 2014), recent research suggests that production in between-subjects recall paradigms interacts with serial position such that a positive production effect emerges for items near the end of a list, while a reverse production effect (i.e., silent > aloud) occurs for early items (e.g., Gionet et al., 2022; Saint-Aubin et al., 2021; for a meta-analysis, see Fawcett et al., 2023).

[2] Our initial design was conceived as three experiments, with the first excluding the verbal label, the second providing the verbal label concurrent to the object, and the third providing the verbal label preceding the object. However, due to challenges with data acquisition during the recent pandemic, our sample was smaller for the third experiment than anticipated; because no differences were observed between the second and third experiments, they have been combined here to maximize statistical power for the critical comparison between the label and non-label groups. While a balanced design is necessitated by some conventional approaches (e.g., ANOVA; see, e.g., Shaw & Mitchell-Olds, 1993), multilevel models are robust to substantial inequalities across cells (Clarke, 2008). Thus, we affirm that the unequal number of participants in each condition had little bearing on the analyses reported herein.

response generation. The label condition was included to adjudicate between these possibilities: If production itself benefits visual representations, a production effect should be observed regardless of whether a label was provided.

Method

Participants

Two hundred and eight students from Memorial University participated (60 no labels, 113 labels presented concurrent to the image, 35 labels presented preceding the image).[3] Our initial sample size for the no-label condition was determined based on Fawcett et al. (2016), who applied a similar analytic technique using a different encoding manipulation (i.e., item-method directed forgetting). Whereas they used 35 participants, we roughly doubled this figure to account for noise inherent in running our task online. The remaining experiments were instead posted for the duration of the term and as many participants as possible were gathered within that time frame. Participants were compensated with partial credit toward an eligible undergraduate course. Participants had normal or corrected-to-normal vision and no self-reported issues pertaining to color perception.

Materials and apparatus

Consent forms were completed and demographic information was collected online via the survey tool Qualtrics prior to participants being forwarded to the main experiment at a separate link. The main experiment was programmed in JavaScript using the JsPsych Library (de Leeuw et al., 2023) and participants completed the study in their own time using their own devices to display the experiment. Production studies have been conducted online in this manner in the past without issue (Fawcett & Ozubko, 2016). Study instructions were presented in 16-px Arial font and the cues used to instruct participants to name items aloud or silently were a pair of 150 × 150 px images depicting either an eye (name silently) or a mouth (name aloud).

For our study materials, 80 photographs depicting objects – each 400 × 400 px in size – were selected from those used by Brady et al., (2013; downloaded from http://timbrady.org/resources.html). This stimulus set was created such that each item was categorically distinct and easily nameable (see Brady et al., 2008). These objects were rotated to a random color (selected according to a uniform distribution) along a LAB color wheel. Colors were assigned in this manner on a participant-by-participant basis and the items split randomly into 40 aloud and 40 silent items.

Procedure

Participants were instructed that they would view a series of common objects, each of which they should try to remember in as much visual detail as possible for a later test. On each trial, they were also told to name the object as quickly as possible according to a visual cue preceding the image itself; if the cue was an eye, they were to name the object silently in their head, whereas if it was a mouth, they were to name the object aloud. Participants were explicitly instructed to remember both the color and location of each object, in addition to its name. Participants in the label condition were further instructed that they would be presented with the name of the object either preceding the object (under the visual cue) or concurrent to the object (in the centre of the screen). Figure 1 depicts the study and test procedure for each condition.

Study phase

Each study trial began with the presentation of a fixation stimulus ("+") for 500 ms. This was then followed by the cue image (i.e., an eye or a mouth) for 1,000 ms and finally the object for 4,000 ms, followed by a 500-ms blank screen. For participants assigned to the label condition, the name of the object was also presented either beneath the cue image or in the centre of the screen concurrent to object presentation. Participants assigned to the no-label condition were required to generate the label themselves and were instructed to guess if they were unable to determine the identity of the object (no data were gathered regarding guessing). Objects were presented at a random location on a circle (depicted by a thin black line) surrounding the fixation point. The study phase included a total of 40 aloud and 40 silent trials.

Test phase

Following completion of the study phase, participants were presented with each object a second time and asked to indicate the original color in which it had been studied; regardless of condition, labels were not provided during the test phase. Each trial began with a fixation stimulus ("+") for 500 ms followed by the image surrounded by a white circle with a smaller circle used as a marker. Participants were instructed to move the marker along the larger circle using the computer mouse to change dynamically the color of

[3] A further 18, 34, and nine participants completed the experiment in the no-label, label presented concurrent, and label presented preceding conditions, respectively, but were excluded for failing attentional checks (i.e., misreporting what they were meant to do for each instruction image), reporting off-task behavior (e.g., watching a movie), or responding to either judgment on average in less than ~ 1 s (suggesting barely enough time to interact with the continuous judgment on most trials). Whilst high for an in-person task, these exclusion rates and justification are typical of online studies as detailed in a recent review of the area (Thomas & Clifford, 2017).

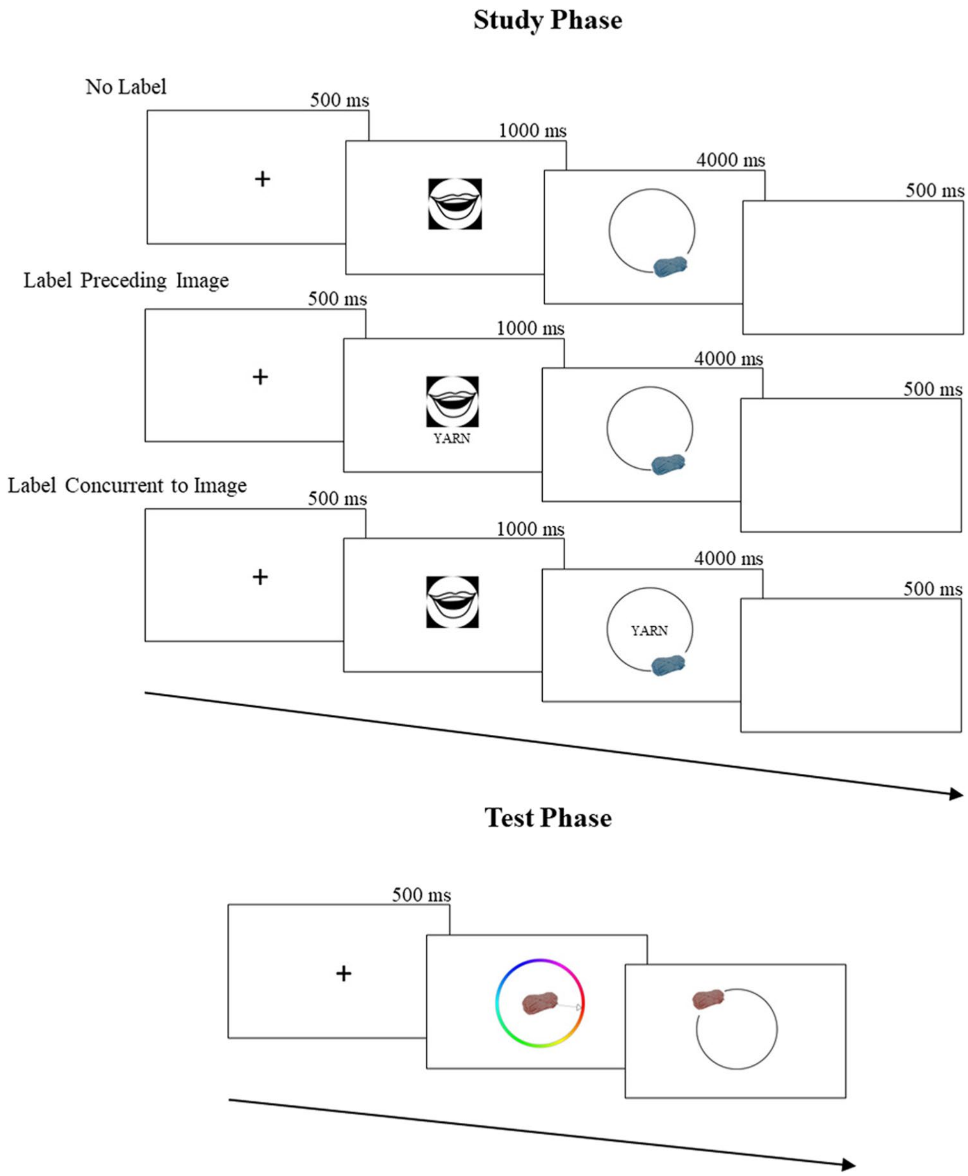

Fig. 1 Study and test phase trial events for all experimental conditions. This figure depicts the procedure used for aloud trials. For silent trials, the mouth icon was replaced by an eye icon and participants instead labeled images covertly; the procedure was otherwise identical. In all conditions, participants were instructed *not* to label images until the image had appeared and the participant had identified the object being depicted

the image. Participants were instructed to move the dongle until the image color matched the original color as closely as possible, submitting their response by clicking the mouse button. Doing so recorded both their response time and the angular distance between the selected and target colors. The object then appeared at a random location on the same circle from the study phase (portrayed in the selected color), and participants moved the location of the image in a manner similar to the color judgment. Responses during this phase were self-paced, but participants were instructed to avoid overthinking their response.

Following the test phase, all participants were redirected to a second Qualtrics survey where they completed a brief questionnaire asking if they followed the instructions from the previous two phases. These questions included asking if they took part in other activities while completing the study, if they wrote down the names of the items or used another strategy besides those instructed, if there was any reason their data should be excluded, and what they were supposed to do when presented with an eye or mouth during the study phase. These questions were used to exclude participants who had technical issues, were distracted, or cheated during the study phase.

Statistical approach

The measure of interest was the angular distance between the target color or location and the color or location selected at test for each image. Both color and location were included as dependent measures for two reasons: First, our intended modelling approach requires a great deal of data – it is often applied in the context of short-term memory, where hundreds of trials might be gathered – and we believed that measuring two dimensions rather than one would provide a more stable estimate. Second, it had been our intent to evaluate on an exploratory basis whether color and location might be differentially impacted by production. For that reason, both measures were analyzed in the same model. An initial model included the dependent measure as an additional independent variable – but failed to observe any difference between the two responses. Accordingly, models reported in-text have collapsed these measures. For interested readers, however, the Online Supplementary Material reports models comparing outcomes for color and spatial position; the inferences that can be derived from these supplemental models are identical to those reported in text. As such, data were analyzed as a function of condition (aloud, silent) using a series of Bayesian multilevel models implemented with the *brms* package (Bürkner, 2017) in the *R 3.5.1* (R Core Team, 2020) programming language. We analyzed our data first using a von Mises model estimating the precision (i.e., the inverse variance) of the color responses; we next fit a variable-precision mixture model designed to estimate separately the probability and fidelity with which the object is represented in memory (e.g., Lawrence, 2010; Zhang & Luck, 2008; for discussion in a similar modelling context, see Fawcett et al., 2016).

The mixture model assumes that each response arises from one of two scenarios: The first is that the participant *cannot* recollect the object color and must therefore guess, producing a deviation sampled from a uniform distribution subtending the circumference of the circle. The second is that the participant *can* recollect the object color, producing a deviation sampled from a von Mises distribution centred on the target color (i.e., 0 radians) with the standard deviation of that distribution reflecting the fidelity with which the object is represented (e.g., greater variability reflects poorer fidelity). By estimating the relative probability of responses being sampled from each distribution, we can estimate the probability of participants recollecting the object's color. Further, by estimating the relative variability of the von Mises distribution (i.e., responses for which the color was recollected), we can estimate the precision with which the color is represented.

For the von Mises model, priors were placed on each condition such that the mean log-transformed precision of each condition could range anywhere between -2 and 2 (roughly 0.14 and 7.4 radians) with individual participant precisions ranging from -4 to 4 (roughly 0.02 to 54.60). These effectively reflect uniform priors, although a mildly regularizing prior was used for the correlations between the random effects. For the precision component of the mixture model, the mean log-transformed precision of each condition could range anywhere between -1 and 3 (roughly 0.37 and 20.10 radians) with individual participants' precisions ranging from -3 to 5 (roughly 0.05 to 148.41 radians). This distribution was shifted in the positive direction slightly, reflecting the fact that the von Mises portion of this model reflected only those trials for which participants had some recollection of the color or location. However, it was still largely uniform. For the recollection component of the mixture model, the mean logit-transformed proportion of each condition could range anywhere between -2.84 and 0.83 (roughly 5.5% and 68.7%) with individual participants estimates ranging from -3.66 to 1.66 (roughly 2.5% and 83.8%). These distributions were slightly biased towards lower recollection values (i.e., not centred on 50%), but were otherwise similarly uniform. A mildly regularizing prior was again used for the correlations between the random effects. Models were fit using 16 independent chains with 2,500 iterations (1,250 warm-up samples) each and convergence was determined via visual inspection as well as standard metrics such as *R*-hat (Gelman & Hill, 2006). This resulted in a total of 40,000 samples, with 20,000 post-warm-up samples.

Results

As depicted in Fig. 2, either modelling approach found production to improve memory for the visual features of the produced items in the no-label condition, as demonstrated by an increase in the probability of remembering the visual feature in question (mixture model) and an increase in the fidelity with which that feature was represented (von Mises and mixture model). Differences and their confidence intervals are provided in Table 1. However, no benefit of production was observed in the label condition.[4]

[4] A comparable Frequentist ANOVA applied to the absolute angular distance between the target and selected feature revealed the same, with a main effect of production, $F(1,206) = 15.67$, $MSe = 41.77$, $p < .001$, and label, $F(1,206) = 13.66$, $MSe = 577.29$, $p < .001$, but also a significant interaction, $F(1,206) = 19.40$, $MSe = 577.29$, $p < .001$, such that a production effect was observed for the no-label group, $t(59) = 5.97$, $p < .001$, but not the label group, $t(147) = 0.97$, $p = .336$. Inclusion of the dependent measure (color, location) as an additional factor produced the same results, with neither the main effect of measure, $F(1,206) = 0.01$, $MSe = 244.62$, $p = .962$, nor any of the interactions involving measure, all $F\text{s} < 1$ and all $p\text{s} > .40$, reached significance.

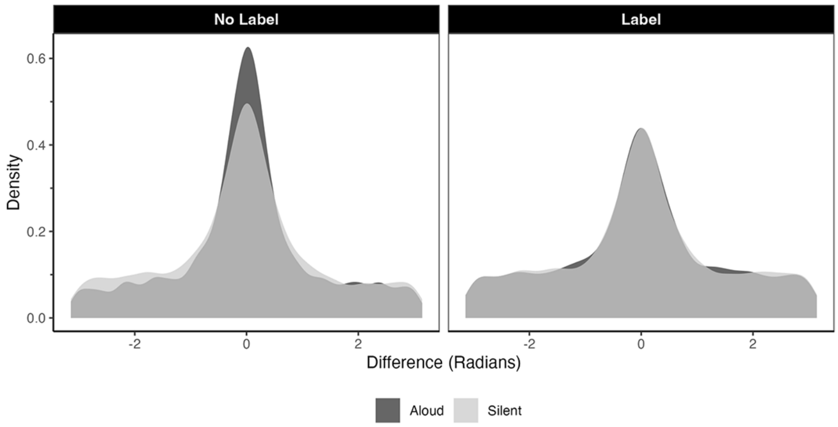

Fig. 2 Density plot of angular error (radians) as a function of production (aloud, silent) and label (label, no label). A value of 0 reflects selection of the precise target color or location

Discussion

The present study is the first to demonstrate that producing the name of a visual object improves the probability of recollecting specific visual elements in long-term memory, as well as improving the fidelity with which those elements are represented, but only when participants self-generated verbal labels; when labels were provided, there was no credible production effect. This finding supports existing work that production also improves performance for tasks thought to be reliant on visual discrimination (Hourihan & Churchill, 2020; Richler et al., 2013). Furthermore, we observed a generation effect (i.e., no-label > label) even for unproduced items (*cf.* Overkott & Souza, 2022; see also, Overkott et al., 2023);[5,6] the probability of remembering visual details was credibly higher in the silent no-label condition relative to both the aloud and silent label conditions, although this benefit did not persist for the precision of those visual details. Interpreted in aggregate, our findings suggest that production can improve memory for visual detail, but they also apply the important caveat that this benefit appears to be dependent on response generation rather than production itself.

Prior to this investigation, two studies reported production effects in memory for tasks reliant on visual discrimination: Both Richler et al. (2013) and Hourihan and Churchill (2020) demonstrated a production effect using forced-choice recognition, wherein participants produced or silently named pictures at study and discriminated between target items and matched exemplars at test (i.e., pictures of items with the same verbal label as the target).[7] In both cases, however, items at study were presented without labels, thereby necessitating that participants self-generate responses (Zormpa et al., 2019a). Given our failure to observe a production effect on memory for visual detail when labels were provided, our findings suggest that benefits observed by Hourihan and Churchill (2020) and Richler et al. (2013) might have been driven by response generation rather than production per se. However, it is important to note that our paradigm differed substantially from earlier investigations: First, our memory test was comprised solely of items previously studied and did not necessitate discrimination between items, thereby eliminating the utility that distinctive visual semantic representations might afford. In forced-choice

[5] Using a paradigm like that of the present study, Overkott and Souza (2022) found that naming unlabelled pictures did not benefit long-term memory for stimulus color. However, participants in Overkott and Souza were subject to tests of short-term memory every three trials and evaluation of long-term memory consisted of 288 trials, more than thrice that of the present study. Given this variation and the fact that long-term memory performance was near floor for participants in Overkott and Souza, we do not believe these results are directly comparable to those of the present study.

[6] More recent efforts by Overkott et al. (2023) demonstrated a mnemonic benefit for color memory in an alternate labeling paradigm, wherein participants labeled the color of an object which was common across trials. Although this paradigm bears similarities to our own investigation, a key feature of the production effect is that the benefit arises only when the productive act is item-specific, distinguishing production from labelling (see, e.g., MacLeod, 2011; MacLeod et al., 2010; Richler et al., 2013). Moreover, Overkott et al. (2023) assessed memory every three trials, which differs substantially from our assessment of participants' long-term memory for a total of 80 items. Given these critical methodological differences, it seems unlikely that the developmental theoretical framework proposed by Overkott et al. (2023) possesses relevant implications for our findings.

[7] Richler et al., (2013, Experiment 2) found that relative to silent reading, producing the names of unlabeled pictures yielded a substantial numerical advantage in forced-choice discrimination between targets and similar exemplars. However, this trend was not subject to statistical analysis.

Table 1 The precision (radians) and probability (%) of memory for color or location judgments as a function of model (von mises, mixture) and label (no label, label); values in parentheses reflect 95% confidence intervals and Bayesian *p*-values reflect confidence in the production effect (PE) or in the superiority of the no-label PE over the label PE. See text for more detail

Model	Silent	Aloud	PE	*p*
Von Mises (Precision)				
No Label	1.76 (1.45, 2.25)	2.76 (1.93, 4.22)	1.00 (0.43, 2.06)	1.00
Label	1.40 (1.28, 1.54)	1.41 (1.29, 1.56)	0.01 (-0.06, 0.09)	.66
Difference (NL – L)			0.99 (0.42, 2.06)	1.00
Mixture (Precision)				
No Label	4.51 (3.54, 5.59)	6.17 (5.07, 7.33)	1.65 (0.32, 3.01)	.99
Label	5.57 (4.59, 6.59)	5.69 (4.75, 6.67)	0.13 (-1.10, 1.34)	.59
Difference (NL – L)			1.53 (0.09, 3.02)	.98
Mixture (Probability)				
No Label	44.9 (36.7, 53.1)	53.8 (44.9, 62.2)	8.9 (3.1, 14.6)	.99
Label	32.7 (27.9, 37.6)	32.3 (27.1, 37.7)	-0.4 (-3.9, 3.3)	.42
Difference (NL – L)			9.3 (2.9, 15.6)	.99

exemplar paradigms, on the other hand, participants might leverage distinctive representations to guide discrimination. Additionally, we evaluated memory only for two specific visual features (i.e., color and spatial position), whereas participants might rely on other features (e.g., shape, texture) to guide discrimination between exemplars. Nonetheless, color and spatial position represent rudimentary visual features that are often processed automatically (e.g., Park & Mason, 1982); if production does not improve memory for these basic features, it seems unlikely that a benefit would extend to more complex visual details or representations.

In contrast to our findings, generation typically produces negative (i.e., read > generate) effects on memory for visual detail (e.g., Mulligan et al., 2006; Nieznánski, 2011, 2012). This pattern has been explained by some with reference to an *item-context trade-off*, wherein generation improves item memory at the expense of hindering memory for contextual details (Jurica & Shimamura, 1999; see also, Nieznánski, 2011, 2012); this trade-off is thought to occur due to the additional cognitive effort necessitated by generation relative to reading. Given that both item color and spatial position represent contextual details, such an account would predict *worse* memory for these details in our no-label condition. However, we instead observed a robust benefit for both color and spatial position when participants self-generated verbal labels, thereby providing evidence against a trade-off account.

On the other hand, the *processing account* provides a more nuanced view of the relationship between generation and context memory. According to this view, generation (relative to reading) elicits a higher degree of semantic, nonvisual processing, allocating cognitive resources away from processing perceptual attributes of the target stimulus (Jacoby, 1983; Mulligan, 2004, 2011); details extraneous to the target are unaffected by differential processing, leading to a null generation effect for extrinsic contextual details (e.g., spatial position) and a negative effect (i.e., read > generate) for intrinsic details (e.g., stimulus color). Critically, however, Mulligan and colleagues (2004, 2011; Mulligan et al., 2006) suggest that generation-related differences in processing are task-specific: While typical generation tasks rely predominantly on nonvisual processing (e.g., antonym generation) and hinder memory for context, tasks that rely instead on visual processing (e.g., letter transposition) should improve memory for visual details. Given that participants in the present study were provided only visual information (i.e., pictures) as a cue to generate the target response, our paradigm necessitated visual processing almost exclusively. Thus, our observation of a positive generation effect on both extrinsic (i.e., location) and intrinsic (i.e., color) visual details provides strong support for the processing account over an item-context trade-off. Although earlier research has observed positive generation effects on memory for extrinsic visual details (e.g., Mulligan et al., 2006), the present study is the first to show that visually driven generation tasks can benefit memory for intrinsic details, validating a key prediction made by the processing account that has hitherto remained unsupported (Mulligan, 2011).

Although our findings align well with processing accounts of the generation effect, it is more difficult to reconcile our observation of a production effect in the no-label condition with theoretical frameworks of production. Critically, our paradigm obviates any utility that relative distinctiveness might afford, given that participants were not required to discriminate between items at test. One possible explanation for this finding is that participants failed to adequately generate labels for items unless an overt response was required. Were this the case, however, we would expect similar performance between the silent no-label and label conditions, but we instead observed an advantage for the silent no-label condition relative to either label condition.[8] Nonetheless, this pattern of results could be accommodated if participants failed to engage deeply with trials that did not require overt generation. With respect to this explanation, it may be the case that participants engaged with silent no-label trials only insofar as pictures were recognized; recognition of common objects is thought to be an automatic cognitive process (e.g., Boucart et al., 2000; Dell'acqua & Job, 1998) and accordingly, activation of semantic labels for these trials may not have necessitated deep visual processing. Congruent with such an explanation, the advantage we observed for the silent no-label condition relative to either label condition occurred only for probability and did not extend to fidelity, which may imply that the benefit was driven by a shallow form of visual processing when generation was covert. On the other hand, trials that required overt generation might have forced participants to carefully consider pictures while generating to-be-produced labels, necessitating deeper engagement with visual features relative to that afforded by automatic processes related to object recognition.

Another possibility is that production *does* enhance memory for visual detail, but that response generation is a prerequisite for this benefit. Distinctiveness-based frameworks cannot readily accommodate a production-related advantage for visual details given that typical accounts would predict the production effect to arise solely on the basis of distinctive articulatory and auditory features (e.g., Forrin et al., 2012; Mama & Icht, 2016), which should possess no additional information that can be used to guide retrieval of visual details. However, the benefit we observed might be accommodated if one accepts that production alters the way participants engage with study items. To contextualize our results, production of self-generated verbal labels might have increased attention to visual details present at encoding; this could result in these details being better encoded, yielding a mnemonic benefit at test. At present, a growing body of evidence is congruent with the hypothesis that production increases attentional engagement at encoding (e.g., Bailey et al., 2021; Hassall et al., 2016; Mama et al., 2018). With respect to the fact that the production effect we observed was confined to the no-label condition, one possibility is that participants attended preferentially to orthographic labels rather than pictures when the former were provided. Relative to reading words, processing images is slower and more effortful (Johnson et al., 1996), and accordingly, the latter may have been prioritized to conserve effort; if participants ignored pictures when labels were present, a production effect would be expected only in the absence of labels. However, this hypothesis is incongruent with our observation that performance in the label condition was well above chance, indicating that participants *did* attend to pictorial stimuli even when labels were present. Furthermore, presenting labels prior to pictures – and thereby eliminating the possibility that orthographic processing would be prioritized – still did not result in a production effect.

Alternatively, a production effect confined to the no-label condition might be explained if production-related attentional enhancements reflect increased engagement solely with the target item itself (i.e., the to-be-produced stimulus) rather than with the multitude of features present in the spatiotemporal environment at encoding (i.e., the study trial). In picture-naming trials for which no label was provided, then, increased engagement with the target item would reflect increased engagement with the pictorial stimulus and, by extension, the visual features of the stimulus; when orthographic labels are provided, however, production increases engagement solely with labels themselves, rather than the visual features of pictorial stimuli present at encoding. Although such an explanation fits neatly with the pattern of results we observed, this hypothesis conflicts with recent findings that production itself does not improve memory for contextual details intrinsic to the target item (e.g., font size; Bodner et al., 2020). Further, in contrast to our findings, this hypothesis would predict a production effect to occur when labels are presented preceding pictures: Orthographic labels disappeared before participants produced the items and accordingly, a production-related attentional increase should benefit memory for pictorial stimuli present during the productive act.

In sum, our findings suggest that improvements to the probability and fidelity of memory for color and spatial position are driven predominantly by response generation rather than production. When generation tasks rely on visual processing, participants must leverage visual details to generate targets from cues, enhancing processing of these details and leading to a mnemonic benefit. On the other hand, production appears only to improve memory for visual detail when generation is a prerequisite for the productive act. While we cannot yet determine whether this benefit is driven by deeper visual processing, attentional increases, or a yet-to-be identified alternative mechanism, evidence appears to favor the former.

[8] For the silent no-label condition, the mixture model revealed mnemonic advantages for the probability of remembering features of $M = 12.1\%$, 95% CI [0.03 – 0.21], and $M = 12.6\%$, 95% CI [0.03 – 0.22], relative to the silent and aloud label conditions, respectively.

Supplementary information The online version contains supplementary material available at https://doi.org/10.3758/s13423-024-02566-5.

Funding Jonathan M. Fawcett was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

Data availability Data and analysis scripts are available at https://doi.org/10.5281/zenodo.13270669. Pictorial stimuli used in the study are available at http://timbrady.org/resources.html.

Code availability Data and analysis scripts are available at https://doi.org/https://doi.org/10.5281/zenodo.13270669.

Declarations

Ethics approval The study was approved by the Interdisciplinary Committee on Ethics in Human Research (ICEHR) at Memorial University of Newfoundland and Labrador. All aspects of the study adhered to ethical standards outlined in the 1964 Helinski Declaration.

Consent to participate All participants in this study provided informed consent to participate.

Consent to publish All participants in this study provided informed consent to have their data published.

Open practices statement The experiment reported herein was not preregistered. The materials for the experiment are available at http://timbrady.org/resources.html. Data and analysis scripts are available at https://doi.org/https://doi.org/10.5281/zenodo.13270669.

Conflicts of interest The authors declare no competing financial or proprietary interests relevant to the content of the article.

References

Bailey, L. M., Bodner, G. E., Matheson, H. E., Stewart, B. M., Roddick, K., O'Neil, K., Simmons, M., Lambert, A. M., Krigolson, O. E., Newman, A. J., & Fawcett, J. M. (2021). Neural correlates of the production effect: An fMRI study. *Brain and Cognition, 152*(2021), 105757. https://doi.org/10.1016/j.bandc.2021.105757

Bellezza, F. S. (1981). Mnemonic devices: Classification, characteristics, and criteria. *Review of Educational Research, 51*(2), 247–275. https://doi.org/10.3102/00346543051002247

Bodner, G. E., Huff, M. J., & Taikh, A. (2020). Pure-list production improves item recognition and sometimes also improves source memory. *Memory & Cognition, 48*, 1281–1294. https://doi.org/10.3758/s13421-020-01044-2

Boucart, M., Meyer, M. E., Pins, D., Humphreys, G. W., Scheiber, C., Gounod, D., & Foucher, J. (2000). Automatic object identification: An fMRI study. *NeuroReport, 11*(11), 2379–2383. https://doi.org/10.1097/00001756-200008030-00009

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 105*(38), 14325–14329. https://doi.org/10.1073/pnas.0803390105

Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science, 24*(6), 981–990. https://doi.org/10.1177/0956797612465439

Bürkner, P. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Clarke, P. (2008). When can group level clustering be ignored? Multilevel models versus single-level models with sparse data. *Journal of Epidemiology and Community Health, 62*(8), 752–761. https://doi.org/10.1136/jech.2007.060798

Conway, M. A., & Gathercole, S. E. (1987). Modality and long-term memory. *Journal of Memory and Language, 26*(3), 341–361. https://doi.org/10.1016/0749-596X(87)90118-5

de Leeuw, J. R., Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software*, *8*(85), 5351. https://doi.org/10.21105/joss.05351

Dell'acqua, R., & Job, R. (1998). Is object recognition automatic? *Psychonomic Bulletin & Review, 5*, 496–503. https://doi.org/10.3758/BF03208828

Ekstrand, B. R., Wallace, W. P., & Underwood, B. J. (1966). Frequency theory of verbal-discrimination learning. *Psychological Review, 73*(6), 566–578. https://doi.org/10.1037/h0023876

Fawcett, J. M. (2013). The production effect benefits performance in between-subject designs: A meta-analysis. *Acta Psychologica, 142*, 1–5. https://doi.org/10.1016/j.actpsy.2012.10.001

Fawcett, J. M., Baldwin, M. M., Whitridge, J. W., Swab, M., Malayang, K., Hiscock, B., Drakes, D. H., & Willoughby, H. V. (2023). Production improves recognition and reduces intrusions in between-subject designs: An updated meta-analysis. *Canadian Journal of Experimental Psychology, 77*, 35–44. https://doi.org/10.1037/cep0000302

Fawcett, J. M., Bodner, G. E., Paulewicz, B., Rose, J., & Wakeham-Lewis, R. (2022). Production can enhance semantic encoding: Evidence from forced-choice recognition with homophone versus synonym lures. *Psychonomic Bulletin & Review, 29*(6), 2256–2263. https://doi.org/10.3758/s13423-022-02140-x

Fawcett, J. M., Lawrence, M. A., & Taylor, T. L. (2016). The representational consequences of intentional forgetting: Impairments to both the probability and fidelity of long-term memory. *Journal of Experimental Psychology: General, 145*, 56–81. https://doi.org/10.1037/xge0000128

Fawcett, J. M., & Ozubko, J. D. (2016). Familiarity, but not recollection, supports the between-subject production effect in recognition memory. *Canadian Journal of Experimental Psychology, 70*(2), 99–115. https://doi.org/10.1037/cep0000089

Fawcett, J. M., Quinlan, C. K., & Taylor, T. L. (2012). Interplay of the production and picture superiority effects: A signal detection analysis. *Memory (hove), 20*(7), 655–666. https://doi.org/10.1080/09658211.2012.693510

Forrin, N. D., MacLeod, C. M., & Ozubko, J. D. (2012). Widening the boundaries of the production effect. *Memory & Cognition, 40*, 1046–1055. https://doi.org/10.3758/s13421-012-0210-8

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gionet, S., Guitard, D., & Saint-Aubin, J. (2022). The production effect interacts with serial positions: Further evidence from a between-subjects manipulation. *Experimental Psychology, 69*, 12–22. https://doi.org/10.1027/1618-3169/a000540

Hassall, C. D., Quinlan, C. K., Turk, D. J., Taylor, T. L., & Krigolson, O. E. (2016). A preliminary investigation into the neural basis of the production effect. *Canadian Journal of Experimental Psychology, 70*(2), 139–146. https://doi.org/10.1037/cep0000093

Hebb, D. O. (1949). *The organization of behavior; a neuropsychological theory*. Wiley.

Hopkins, R. H., & Edwards, R. E. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning and Verbal Behavior, 11*(4), 534–537. https://doi.org/10.1016/S0022-5371(72)80036-7

Hourihan, K. L., & Churchill, L. A. (2020). Production of picture names improves picture recognition. *Canadian Journal of Experimental Psychology, 74*, 35–43. https://doi.org/10.1037/cep0000185

Jacoby, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior, 22*(5), 485–508. https://doi.org/10.1016/S0022-5371(83)90301-8

James, W. (1890). *The principles of psychology, Vol. 1.* Henry Holt and Co. https://doi.org/10.1037/10538-000

Johnson, C. J., Paivio, A., & Clark, J. M. (1996). Cognitive components of picture naming. *Psychological Bulletin, 120*, 113–139. https://doi.org/10.1037/0033-2909.120.1.113

Jones, A. C., & Pyc, M. A. (2014). The production effect: Costs and benefits in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 300–305. https://doi.org/10.1037/a0033337

Jurica, P. J., & Shimamura, A. P. (1999). Monitoring item and source information: Evidence for a negative generation effect in source memory. *Memory & Cognition, 27*(4), 648–656. https://doi.org/10.3758/BF03211558

Lawrence, M. A. (2010). Estimating the probability and fidelity of memory. *Behavior Research Methods, 42*(4), 957–968. https://doi.org/10.3758/BRM.42.4.957

Lin, O. Y. H., & MacLeod, C. M. (2012). Aging and the production effect: A test of the distinctiveness account. *Canadian Journal of Experimental Psychology, 66*(3), 212–216. https://doi.org/10.1037/a0028309

MacLeod, C. M. (2011). I said, you said: The production effect gets personal. *Psychonomic Bulletin & Review, 18*(6), 1197–1202. https://doi.org/10.3758/s13423-011-0168-8

MacLeod, C. M., & Bodner, G. E. (2017). The production effect in memory. *Current Directions in Psychological Science, 26*(4), 390–395. https://doi.org/10.1177/0963721417691356

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. https://doi.org/10.1037/a0018785

MacLeod, C. M., Ozubko, J. D., Hourihan, K. L., & Major, J. C. (2022). The production effect is consistent over material variations: Support for the distinctiveness account. *Memory, 30*(8), 1000–1007. https://doi.org/10.1080/09658211.2022.2069270

Mama, Y., Fostick, L., & Icht, M. (2018). The impact of different background noises on the production effect. *Acta Psychologica, 185*, 235–242. https://doi.org/10.1016/j.actpsy.2018.03.002

Mama, Y., & Icht, M. (2016). Auditioning the distinctiveness account: Expanding the production effect to the auditory modality reveals the superiority of writing over vocalising. *Memory, 24*(1), 98–113. https://doi.org/10.1080/09658211.2014.986135

Mulligan, N. W. (2004). Generation and Memory for Contextual Detail. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(4), 838–855. https://doi.org/10.1037/0278-7393.30.4.838

Mulligan, N. W. (2011). Generation disrupts memory for intrinsic context but not extrinsic context. *The Quarterly Journal of Experimental Psychology, 64*(8), 1543–1562. https://doi.org/10.1080/17470218.2011.562980

Mulligan, N. W., Lozito, J. P., & Rosner, Z. A. (2006). Generation and context memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(4), 836–846. https://doi.org/10.1037/0278-7393.32.4.836

Nieznański, M. (2011). Generation difficulty and memory for source. *The Quarterly Journal of Experimental Psychology, 64*(8), 1593–1608. https://doi.org/10.1080/17470218.2011.566620

Nieznański, M. (2012). Effects of generation on source memory: A test of the resource tradeoff versus processing hypothesis. *Journal of Cognitive Psychology, 24*(7), 765–780. https://doi.org/10.1080/20445911.2012.690555

Overkott, C., & Souza, A. S. (2022). Verbal descriptions improve visual working memory but have limited impact on visual long-term memory. *Journal of Experimental Psychology: General, 151*(2), 321–347. https://doi.org/10.1037/xge0001084

Overkott, C., Souza, A. S., & Morey, C. C. (2023). The developing impact of verbal labels on visual memories in children. *Journal of Experimental Psychology: General, 152*(3), 825–838. https://doi.org/10.1037/xge0001305

Ozubko, J. D., Gopie, N., & MacLeod, C. M. (2012). Production benefits both recollection and familiarity. *Memory & Cognition, 40*(3), 326–338. https://doi.org/10.3758/s13421-011-0165-1

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*(6), 1543–1547. https://doi.org/10.1037/a0020604

Ozubko, J. D., Major, J., & MacLeod, C. M. (2014). Remembered study mode: Support for the distinctiveness account of the production effect. *Memory, 22*(5), 509–524. https://doi.org/10.1080/09658211.2013.800554

Park, D. C., & Mason, D. A. (1982). Is there evidence for automatic processing of spatial and color attributes present in pictures and words? *Memory & Cognition, 10*, 76–81. https://doi.org/10.3758/BF03197628

Quinlan, C. K., & Taylor, T. L. (2013). Enhancing the production effect in memory. *Memory (hove), 21*(8), 904–915. https://doi.org/10.1080/09658211.2013.766754

Quinlan, C. K., & Taylor, T. L. (2019). Mechanisms Underlying the Production Effect for Singing. *Canadian Journal of Experimental Psychology, 73*(4), 254–264. https://doi.org/10.1037/cep0000179

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/. Accessed 1 Jan 2024

Richler, J. J., Palmeri, T. J., & Gauthier, I. (2013). How does using object names influence visual recognition memory? *Journal of Memory and Language, 68*, 10–25. https://doi.org/10.1016/j.jml.2012.09.001

Saint-Aubin, J., Yearsley, J. M., Poirier, M., Cyr, V., & Guitard, D. (2021). A model of the production effect over the short-term: The cost of relative distinctiveness. *Journal of Memory and Language, 118*, 104219. https://doi.org/10.1016/j.jml.2021.104219

Thomas, K. A., & Clifford, S. (2017). Validity and Mechanical Turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior, 77*, 184–197. https://doi.org/10.1016/j.chb.2017.08.038

Varao Sousa, T. L., Carriere, J. S., & Smilek, D. (2013). The way we encounter reading material influences how frequently we mind wander. *Frontiers in Psychology, 4*, 892. https://doi.org/10.3389/fpsyg.2013.00892

Willoughby, H. V., Tiller, J., Hourihan, K. L., & Fawcett, J. M. (2019). *The pupillometric production effect: Measuring attentional engagement during a production task* [Paper presentation]. CSBBCS 2019 Meeting, Waterloo, Canada.

Zhang, B., Meng, Z., Li, Q., Chen, A., & Bodner, G. E. (2023). EEG-based univariate and multivariate analyses reveal that multiple processes contribute to the production effect in recognition. *Cortex, 165*, 57–69. https://doi.org/10.1016/j.cortex.2023.04.006

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*(7192), 233–235. https://doi.org/10.1038/nature06860

Zormpa, E., Brehm, L. E., Hoedemaker, R. S., & Meyer, A. S. (2019a). The production effect and the generation effect improve memory in picture naming. *Memory, 27*(3), 340–352. https://doi.org/10.1080/09658211.2018.1510966

Zormpa, E., Meyer, A. S., & Brehm, L. E. (2019b). Slow naming of pictures facilitates memory for their names. *Psychonomic Bulletin & Review, 26*, 1675–1682. https://doi.org/10.3758/s13423-019-01620-x