This version of the manuscript has been accepted for publication. The original article can be found under

https://doi.org/10.1177/0956797620965519

Cite as:

Ovalle-Fresa, R. Uslu, A.S. & Rothen, N. (2021). Levels of processing affect perceptual features in visual-associative memory, *Psychological Science 32*(2), 267-279. https://doi.org/10.1177/0956797620965519

Levels of processing affect perceptual features in visual-associative memory

Rebecca Ovalle-Fresa[1,2], Arif Sinan Uslu[1,3] and Nicolas Rothen[1,2]

[1] Institute of Psychology, University of Bern, Switzerland

[2] Faculty of Psychology, Swiss Distance University Institute, Switzerland

[3] Institute for Health and Behaviour, University of Luxembourg, Luxembourg

Author Note

Correspondence concerning this article should be addressed to Rebecca Ovalle Fresa, University of Bern, Institute of Psychology, Fabrikstrasse 8, 3012 Bern, Switzerland. E-mail: rebecca.ovallefresa@gmail.com

This work was supported by the Swiss National Science Foundation (Grant Number: PZ00P1_154954).

Abstract

The levels of processing (LOP) account has inspired thousands of studies with verbal material. The few studies investigating LOP with non-verbal stimuli used images with nameable objects which, like meaningful words, lend themselves to semantic processing. Thus, nothing is known about effects of different LOP on basic visual perceptual features, such as color. Across four experiments, we tested 187 participants to investigate whether the LOP framework also applies to basic perceptual features in visual-associative memory. For Experiments 1 and 2, we developed a paradigm to investigate recognition memory for associations of basic visual features. Participants had to memorize object-color associations (Exp. 1) and fractal-color associations (Exp. 2, to suppress verbalization). In Experiments 3 and 4, we extended our account to cued recall. All experiments revealed reliable LOP effects for basic perceptual features in visual-associative memory. Our findings demonstrate that the LOP account is more universal than the current literature suggests.

Statement of Relevance

Long-term memory is usually assumed to depend critically on meaning. The more meaningfully stimuli can be processed during encoding, the more likely they can be retrieved later. This notion is reflected in the levels of processing account (LOP). Over the last 50 years, LOP has been studied almost exclusively with verbal material and it is not known whether it also applies to basic visual perceptual features (e.g., color). In four experiments, we tested LOP on perceptual features in visual-associative memory. We obtained evidence that LOP also applies to visual-associative memory. This was not only

the case for associations between concrete objects and colors but also between abstract fractals and colors, even though the stimuli were very hard to verbalize in the latter case. This provides evidence that the fundamental principle of LOP also applies to visual mental representations and that it is much more universal than the current literature suggests.

Keywords: Levels of Processing, Abstract Stimuli, Memory, Perception, Color

Levels of processing affect perceptual features in visual-associative memory

The levels of processing (LOP) account denotes that deeper processing during encoding (e.g., semantic) is beneficial for subsequent memory retrieval, compared to superficial processing (e.g., perceptual; Craik & Lockhart, 1972). Over the last 50 years, it has inspired thousands of studies with verbal material (e.g., visually presented words). Despite the importance of memory based on visual impressions in everyday life, studies investigating LOP with non-verbal material (e.g., visual images) are remarkably scarce (e.g., Baddeley & Hitch, 2017; D'Agostino, O'Neill, & Paivio, 1977; Intraub & Nicklos, 1985; Xu, Zhao, Zhao, & Yang, 2011). Even more, nothing is known about the effects of different processing levels on basic visual perceptual features, such as color. Here, we present an innovative approach – using shape-color associations – and assess in a series of four experiments if the LOP account also applies to basic visual perceptual features.

Extending the findings from the LOP literature to basic visual perceptual features bears the potential to provide important insights into the extent and universality of this important framework. Foremost, it will inform us whether the LOP framework only applies to verbal stimuli, as the lack of studies on LOP effects with visual stimuli might suggest or whether the LOP framework is much more universal than several decades of empirical evidence are currently suggesting. That is, there is general agreement in the LOP literature that deeper processing is concerned with semantic processing (Craik & Lockhart, 1972). Unsurprisingly, research has almost exclusively focused on stimuli which are easy to verbalize, even when visual stimuli were used (e.g., Baddeley & Hitch, 2017; D'Agostino, O'Neill, & Paivio, 1977; Intraub & Nicklos, 1985; Xu, Zhao, Zhao, & Yang, 2011). LOP effects on basic visual perceptual features would provide strong evidence that the LOP framework is much more universal with regards to human cognition than currently embraced by cognitive scientists and psychologists more generally.

To corroborate the effects of LOP manipulations on basic perceptual features in visual memory, we conducted four memory experiments based on two different paradigms with monochrome colors as an integrative feature of different shapes. In all experiments, deep encoding required a pleasantness judgement, shallow encoding a judgement on the presence of straight lines. In Experiment 1, we investigated LOP effects on recognition memory for object-color associations (concrete shapes). In Experiment 2, we applied the same paradigm on fractal-color associations (abstract shapes), to suppress verbalization. Experiment 1 and 2 provide insight if a shape-color association, particularly the perceptual features, in memory representations are accessed or not (i.e., if LOP affect the quantity of accessed representations). In Experiment 3 and 4, we extended our approach, and investigated LOP effects in a recall task with a continuous response space. Here, we took advantage of the continuous nature of the dependent measure and applied a computational modeling approach (Brady, Konkle, Gill, Oliva, & Alvarez, 2013; Zhang & Luck, 2008) to estimate whether potential LOP effects are based on access to memory or enhanced quality of memory representations. In Experiment 3, we used object-color and in Experiment 4 fractal-color associations, respectively. Experiment 3 and 4 provide insight if LOP affects not only the quantity of accessed memory representations, but also their quality. Note that this is not possible with verbal / categorical stimuli alone, but only with basic visual perceptual features that are based on continuous representations. Consequently, this is something which has not been addressed by previous research on LOP effects and will thus critically contribute to the literature on the LOP framework. Overall, the continuous nature of the response measure has the advantage that the magnitude of LOP effects can be estimated on a more fine-grained level than previously attempted in empirical studies on LOP effects.

To anticipate our findings, we observed extremely reliable LOP effects in all our experiments. Consistent with the notion that abstract fractals are hard to verbalize, performance was consistently lower for fractal-color associations (Exp. 2 and 4) in comparison to object-color associations (Exp. 1 and 3). Based on computational modeling (Exp. 3 and 4), LOP effects for basic visual perceptual stimuli are based on access to memory, rather than enhanced precision of memory representations. Overall, our findings demonstrate that the LOP account is significantly more universal than previous research implies.

Collectively, our results provide strong evidence that LOP effects can also be observed for basic visual perceptual features. That is, LOP effects are much more universal than previously thought.

Experiment 1

Methods

Participants. Sample size determination was based on the observation of medium sized LOP effects for images (Cohen’s $d = 0.50$; Baddeley & Hitch, 2017). Estimated on the basis of the results for door images across experiments in Baddeley and Hitch (2017), we used $M = .75$ for the deep conditions and $M = .65$ for the shallow conditions and the common $SD = .3$ to perform the a priori power analysis with Superpower (https://arcstats.io/shiny/anova-exact/) (Lakens & Caldwell, 2019). Moreover, we assumed a medium average correlation (*r* = .50) and aimed for a power level of 0.95. The calculations included the within subject factors LOP (shallow vs. deep) and Lure (shape, color, control). It revealed that 40 participants were required to achieve a 95% probability of detecting a true main effect for LOP.

We tested 40 participants with an average age of 21 years (*SD* = 3 years). Eight participants were men, 5 participants were left-handed. All participants were psychology students of the University of Bern who earned course credits in return for participation. Inclusion criteria was normal (or corrected to normal) vision. The study was approved by the local ethics committee of the University of Bern. Participants were informed that they could withdraw at any time during the experiment before they consented to participate.

Apparatus. The tasks were presented with the MATLAB Psychophysics Toolbox 3 (Brainard, 1997; Pelli, 1997) on a flat 21-inch Sony Triniton cathode ray tube monitor (85 Hz refresh rate) driven by a NVIDIA GeForce GTX970 graphics card (NVIDIA Corporation, Santa Clara, California) with a spatial resolution of 1280 × 1024 pixels and a color resolution of 8 bits per channel. The monitor was calibrated, and the stimulus materials rendered to keep the averaged luminance of individual stimuli during the task isoluminant with the background of the monitor (40.7 cd/m2). A more detailed description of monitor calibration and stimulus rendering is provided in Ovalle Fresa and & Rothen (2019). To ensure a constant viewing distance of 100 cm, we used a chin- and forehead rest. To shield peripheral vision and to ensure that the participants adapted to the computer monitor, we installed a black cardboard tunnel from the monitor to the chin- and forehead rest, overlapping the rest by 20 cm at the top and both sides.

Design. The recognition task followed a 2 x 3 within subject design, manipulating *LOP* (shallow vs. deep) and *Lure* (color, object, control). The entire task consisted of one block per experimental condition (total of six blocks). Each block consisted of an encoding and a recognition phase. Each LOP condition consisted of three blocks (i.e., one per lure condition).

Material. Stimuli consisted of a total of 192 distinct and identifiable objects from the Bank of Standardized Stimuli (Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010) and a total of 96 colors equidistantly sampled on an isoluminant color wheel in CIELUV color space. Table 1 shows how the stimuli were assigned to the different conditions. Objects were randomly assigned to two sets (A and B). The objects of one set (A) were randomly assigned to six lists of 16 objects (i.e., one list per LOP and lure condition). The objects of the other set (B) were randomly assigned to two sets of 48 objects, to serve as lures for the control condition. Sets and lists were counterbalanced between participants. An additional set of 16 objects served as practice list. Colors were assigned to six lists. Each color list was associated with one of the object lists. Colors within each list were as distinct from each other as possible. An additional set of four colors served as practice list.

Table 1. *Stimulus lists. Two sets of objects and one set of colors were assigned to the different task conditions. One set of objects was used to for the different lure conditions (color, object, control) and the objects of the other set were used as novel stimuli in the control condition (cf. Materials section). The sets were counterbalanced across conditions.*

	Sets		
	Object		Color
	A	B	
Total	96	96	96
Shallow			
color	16		16
object	16		16
control	16	48	16
Deep			
color	16		16
object	16		16
control	16	48	16

Procedure. Participants were individually tested. First, we tested visual acuity (Landolt ring test; www.sov.ch; cf. Wesemann, Schiefer, & Bach, 2010) and assessed demographic data. Then the room was completely darkened for the rest of the experiment, with the monitor as the only source of light. Afterwards, we assessed color-vision with an online version of the Farnsworth Dichotomous Test (D-15; www.color-blindness.com/color-arrangement-test). Thereafter, participants completed the recognition task, including LOP manipulations. Next, we conducted a contrast sensitivity task which is not relevant for the purpose of the current study and hence not further reported. At the end of the session, a post-test interview was conducted to assess potential strategies and subjective task difficulty.

The recognition task consisted of the two LOP conditions (shallow vs. deep). Each LOP condition was further divided in three lure conditions (shape, color, control).

Order of LOP conditions was counterbalanced between participants. Order of lure conditions within LOP conditions was also counterbalanced between participants. Each of the six conditions consisted of a practice block with four encoding trials, immediately followed by four recognition trials. The practice block was followed by the experimental block, including an encoding phase consisting of 16 trials and an immediately following recognition phase of 16 trials. During encoding, participants were instructed to memorize the object-color associations for a subsequent memory test. Object-color associations were randomized between participants. The order of object presentation during encoding and recognition was randomized for each participant.

Each encoding trial (Figure 1a, c) started with a central fixation (1 s), followed by a colored object (1 s) and the respective encoding question. Deep encoding required a pleasantness judgement ("Y" = pleasant, "N" = not pleasant). Shallow encoding required a perceptual judgement ("Y" = object includes a straight line, "N" = object does not include a straight line). The next trial was initiated by a response or after 2 seconds elapsed. Responses were given on a QWERTZ keyboard. Note that on the QWERTZ keyboard, the "Y" and the "Z" are on interchanged positions, compared to the QWERTY keyboard commonly used in the US and UK. The recognition phase followed the logic of a four-alternative-forced-choice task (4-AFC). During each trial, participants were instructed to choose which of four object-color associations was identical to an object-color association from the encoding phase (i.e., target). A recognition trial (Figure 1b, d) started with a central fixation (1 s) followed by a simultaneous presentation of a target and three lures, each on one of four predefined positions (i.e., at 25%, 37.5%, 62.5%, and 75% on the x-axis of the screen) corresponding to the response keys ("Y", "X", "N", "M"). The position for the target and lures was randomized for each trial. The next trial was initiated by a response or after 8 seconds elapsed. The different lure conditions are presented in Figure 1b and explained in the caption.

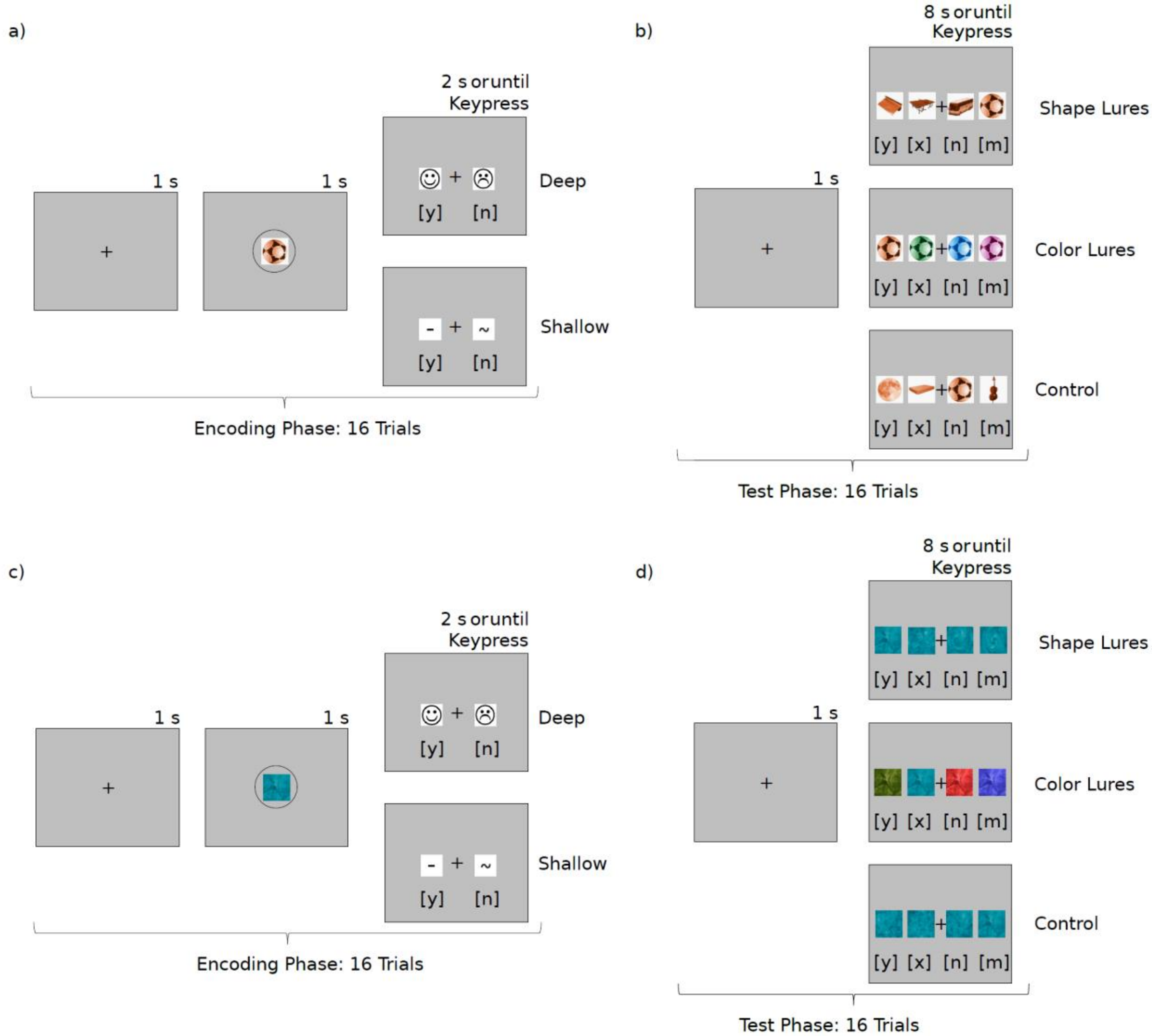

Figure 1. Example trial sequences from the encoding phase (a, c) and recognition phase (b, d) of Experiment 1 (top row) and Experiment 2 (bottom row). Experiments 1 and 2 differed only in whether stimuli were objects or fractals, respectively. In the encoding phase, participants first saw a stimulus and then made either a pleasantness judgment (deep encoding) or a judgment on the presence of straight lines (shallow encoding). In the subsequent recognition phase, participants saw a target and three lures. Participants were

instructed to choose which of four object–color associations was identical to an object–color association from the encoding phase (i.e., target). The lures in the recognition phase involved shapes, colors, or control objects. In the shape-lure condition, the four different shapes had an identical color. Only one of the shapes had been presented in the specific color during encoding (target). The other three shapes (lures) had also been presented during encoding but in a different color. In the color-lure condition, the four shapes were identical but had different colors. All colors were part of the corresponding encoding phase, but three were associated with different shapes from those during encoding (lures), and one was associated with the same shape as during encoding (target). In the control condition, different shapes had the same color, but only one shape had been part of the encoding phase (target); the remaining shapes were novel and had not been part of the encoding phase (lures). The novel objects were taken from Set B (see Table 1) and not otherwise used for a given participant.

Analyses. Raw data are available on Open Science Framework (OSF; DOI 10.17605/OSF.IO/6DE2B). Data analyses were computed in R (R Core Team, 2019). Alpha was set to 0.05 for all analyses. Effect sizes of the analyses of variance (ANOVAs) are reported as partial eta squared (η_p^2). In case of violation of sphericity, Greenhouse Geisser correction was applied. For *t*-tests concerning LOP effects, effect sizes (Cohen's d_z) with 95% confidence intervals were calculated using the R-package rstatix (Kassambara, 2019).

Results

Performance (proportion correct) of Experiment 1 is depicted in Figure 2a. We observed a ceiling effect in the control condition, undermining a reasonable interpretation of this condition. Therefore, the control condition was excluded from all further analysis. We conducted a 2 x 2 ANOVA for repeated measures with the factors *LOP* (shallow vs. deep) and *Lure* (shape vs. color). It revealed a significant main effect *LOP*, $F(1, 39) = 20.39$, $p < .001$, $\eta_p^2 = .34$, based on significant LOP effects for shape lures, $t(39) = 3.02$, $p = .004$, $d_z = 0.41$ [0.18, 0.79], and for color lures, $t(39) = 4.68$, $p < .001$, $d_z = 0.74$ [0.44, 1.10]. However, the main effect *Lure*, $F(1, 39) = 0.00$, $p = .972$, $\eta_p^2 < .01$, and the interaction *LOP* x *Lure* were not significant, $F(1, 39) = 0.89$, $p = .351$, $\eta_p^2 = .02$. Overall, the analysis indicates a clear LOP effect for object-color associations, but no significant influence of different types of lures.

The results from the post-test interview were in line with the behavioral findings: 63% of the participants indicated that the deep LOP condition was easier than the shallow LOP condition. Moreover, 65% indicated that they verbalized the colors and 63% indicated that they verbalized the objects across LOP conditions.

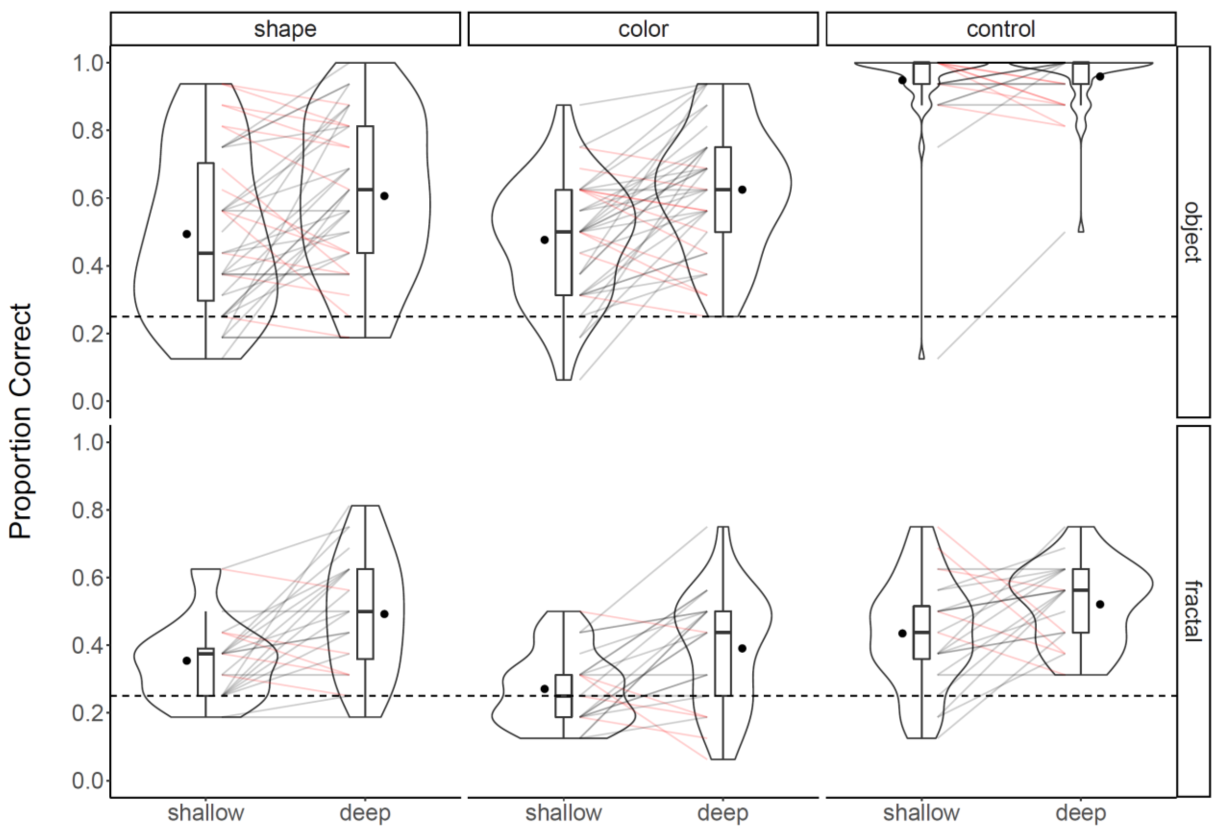

Figure 2. Proportion of correct responses as a function of level of processing (LOP) condition (shallow, deep) and lure (shape, color, control), separately for Experiment 1 (objects; top row) and Experiment 2 (fractals; bottom row). The horizontal line inside each box plot indicates median performance (i.e., 50th percentile), and the dot next to each box plot represents average performance. The lower and upper ends of each box represent the 25th and 75th percentiles, enclosing the interquartile range. The whiskers extend to 1.5 times the interquartile range from either end of the box. The violin plots indicate the density of the data. The colored lines indicate change in performance of

individual participants between the two LOP conditions. Light gray lines represent participants whose performance improved from the shallow to the deep condition. Light red lines represent participants whose performance decreased from the shallow to the deep condition. The dashed line represents the guess rate.

Experiment 2

The primary goal of Experiment 2 was to replicate the LOP effect on the basic visual perceptual feature of color. We ran the same recognition task as in Experiment 1 with abstract fractals instead of concrete objects to suppress verbalization. This puts additional weight on visual perceptual processing in comparison to Experiment 1. Experiment 2 will thus further corroborate whether LOP can affect perceptual features in visual-associative memory.

Methods

Participants. The same a priori power analysis as before revealed that 23 participants were required to replicate the finding with shape lures of Experiment 1 with a 95% probability. We aimed for 24 participants and tested a total of 27 participants, as three had to be replaced due to technical errors (missing data). The average age of the final sample was 21 years (*SD* = 1.81 years), five participants were men and all participants were right-handed.

Material and procedure. Experiment 2 was identical to Experiment 1, with the following exceptions: we used abstract shapes to suppress verbalization. Stimuli consisted of 196 abstract fractals, randomly selected from an initial pool of 400 fractal pictures, originally collected from the internet (search string “fractals”). Fractals of the initial pool were grey-scaled and resized to 380 x 380 pixels, and finally equalized with regards to average spatial frequency and luminance with “The SHINE toolbox”

(Willenbockel et al., 2010). The pool of 400 fractals and their normative values are provided on OSF (DOI 10.17605/OSF.IO/CKFMV). Examples of isoluminant colored fractals can be found on OSF (DOI 10.17605/OSF.IO/6DE2B).

Results

To analyze performance (Figure 2b), data were subjected to a 2 x 3 ANOVA, including the within subject factors *LOP* (shallow vs. deep) and *Lure* (shape, color, control). Raw data are available on OSF (DOI 10.17605/OSF.IO/6DE2B). The ANOVA revealed a significant main effect of *LOP*, $F(1, 23) = 26.34$, $p < .001$, $\eta_p^2 = .53$, indicating better performance after deep encoding than shallow encoding in all conditions (shape: $t(23) = 3.83$, $p = .001$, $d_z = 0.78$ [0.46, 1.25], color: $t(23) = 3.50$, $p = .002$, $d_z = 0.71$ [0.32, 1.31], control: $t(23) = 2.26$, $p = .033$, $d_z = 0.46$, [0.07, 1.09], and *Lure*, $F(1, 46) = 12.95$, $p < .001$, $\eta_p^2 = .36$ (Greenhouse-Geisser correction applied), due to better performance in the control condition than the color condition, $t(23) = 5.82$, $p < .001$, $d_z = 1.19$ [0.76, 1.97], and the shape condition, $t(23) = 2.14$, $p = .044$, $d_z = 0.44$ [0.04, 0.98], and better performance for the shape condition than the color condition $t(23) = 2.60$, $p = .016$, $d_z = 0.53$ [0.15, 0.99]. The interaction was not significant, $F(2, 46) = 0.58$, $p = .566$, $\eta_p^2 = .02$ (Greenhouse-Geisser correction applied). The analyses indicate that medium to large LOP effects are observed in recognition performance for abstract stimuli, but that the LOP effect was not influenced by the type of the lures.

In line with the behavioral results, the post-test interview revealed that more participants (63%) experienced the deep LOP condition as easier compared to the shallow LOP condition. Sixty-five percent of the participants indicated that they verbalized the colors and 61% indicated that they verbalized the fractals during encoding.

Experiment 3

The primary goal of Experiment 3 was to assess the universality of the previous findings (i.e., LOP effects for basic visual perceptual features in recognition memory). Therefore, we investigated whether these LOP effects also apply to basic visual perceptual features in cued recall. We used the same LOP manipulations during encoding as in Experiment 1 and 2. Thereafter, the color of an object had to be recalled and reproduced in a continuous response space, and the difference between the recalled color and the original color at encoding was measured.

Methods

Participants. Due to the implementation of a novel paradigm, sample size determination was again based on the observation of medium sized LOP effects for images (Cohen's *d* = .50; Baddeley & Hitch, 2017) and an a priori power analysis with G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007). The calculations included the within subject comparison *LOP* (shallow vs. deep; two-tailed; alpha = 0.05). We aimed for a power level of 0.95. It revealed that 54 participants were required to achieve a 95% probability of detecting a true main effect for LOP.

We tested a total of 60 participants. Two participants had to be excluded due to incomplete data. The remaining 58 participants had a mean age of 22 years (*SD* = 3.38, min = 19, max = 42), 44 were women, 52 were right-handed.

Apparatus. Hardware and software were the same as used in Experiments 1 and 2. Also, the laboratory setting was identical to Experiments 1 and 2.

Material. Stimuli consisted of 243 unique pictures of objects taken from the BOSS database (Brodeur et al., 2010). Stimulus preparation followed the same procedure as in Experiment 1. All stimuli fitted into 4 x 4 degrees of visual angle and were surrounded by a black circle (representing the color wheel) with a radius of 2.5 degrees of visual angle. The stimuli were randomly assigned to a practice list of three stimuli and two sets of 120 stimuli each, to be used in the shallow or deep condition, respectively. The stimuli of each set were used to create twelve stimulus lists of ten stimuli. Correspondingly, a practice list of three colors and two sets of 120 colors each were created. The colors of both sets were equidistantly sampled on an isoluminant color wheel in CIELUV color space. The colors of the two sets differed by one degree on the color wheel. The colors of each set were further assigned to twelve lists of ten colors each. The colors in each list were always as distinct as possible.

Procedure. The overall procedure was identical to Experiments 1 and 2. In the memory task, we manipulated LOP as in the previous experiments. In the shallow condition, participants were instructed to indicate if a stimulus contained a straight line. In the deep condition, participants were instructed to indicate whether the stimulus was pleasant. The order of the LOP conditions was controlled between participants. The stimulus sets were counterbalanced between LOP conditions. Each condition consisted of 12 blocks with 10 unique objects per block (i.e., objects from one of the stimulus lists). Object-color associations were randomized within each block. Each block consisted of an encoding phase and an immediately following recall phase. During encoding, all ten objects of a given stimulus list were presented in random order. Each object was presented for 3 seconds. During the recall phase, all 10 objects from the preceding encoding phase were presented again in random order, one at a time. Encoding trials were identical to Experiment 1 (Figure 3a). Recall trials (Figure 3b) started with a central fixation cross for one second. Then, an object was presented in grayscale, surrounded by a black circle representing a color wheel. Participants were instructed to adjust the color of a presented object to the original color of the object, as it was presented during encoding. A dot on the color wheel represented to the position of the mouse. Moving the dot with the mouse changed the color of the presented object. A color was selected by mouse click, which also initiated the next trial. If no response was submitted after 8 seconds elapsed, the next trial was initiated. For each recall trial, the following feedback was provided: good, very good and perfect, if the deviation of the response was less or equal to 10, 5 and 0 degrees, respectively.

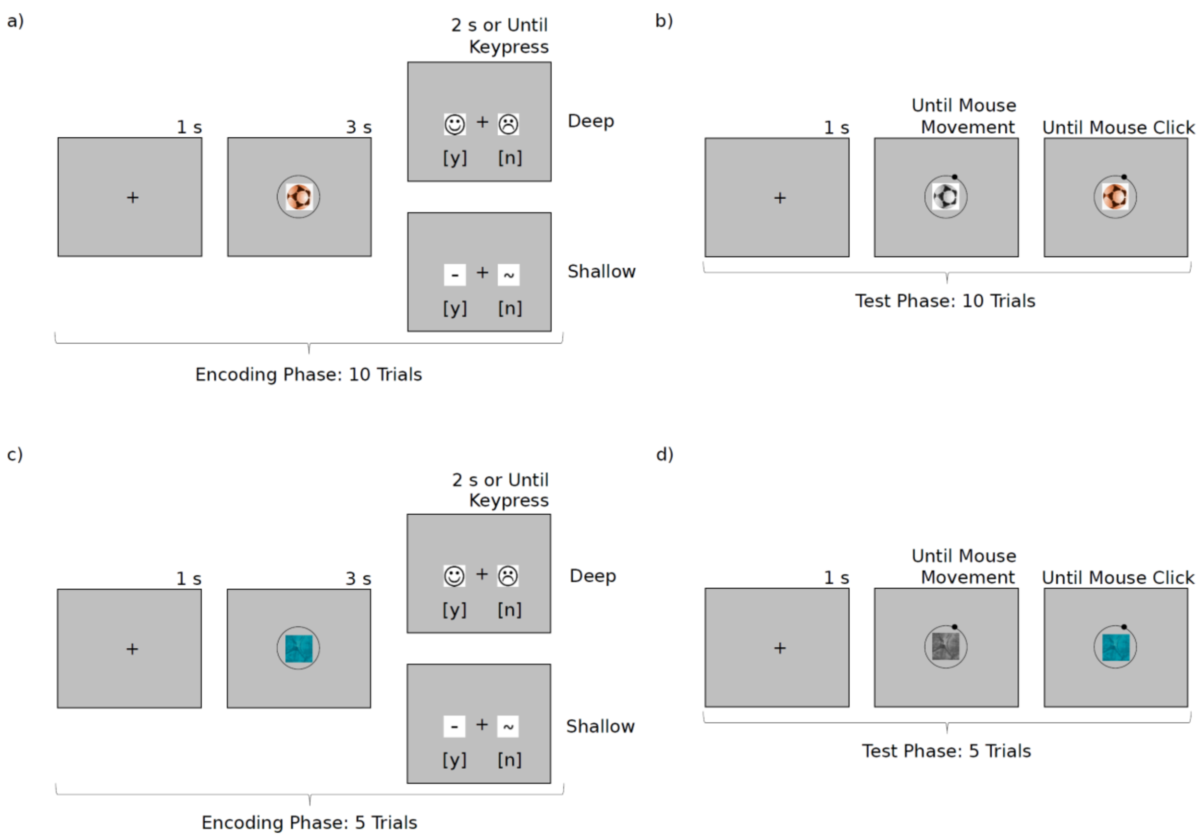

Figure 3. Example trial sequences from the encoding phase (a, c) and recall phase (b, d) of Experiment 3 (top row) and Experiment 4 (bottom row). Experiments 3 and 4 differed only in whether stimuli were objects or fractals, respectively. During encoding, all 10

objects of a given stimulus list were presented in random order, and participants made either a pleasantness judgment (deep encoding) or a judgment on the presence of straight lines (shallow encoding). During the recall phase, each of the 10 objects from the encoding phase was presented again in random order, one at a time. For each object, participants had to adjust a color wheel so the object's color matched how they recalled it originally appearing during encoding.

Design and Analysis. The task followed a within subject design with two conditions (shallow vs. deep). We measured the response error as deviation between original color at encoding and a given response on the color wheel (ranging from 0 to ±180 degrees). Lower values reflect better memory performance. Medians of the absolute response errors were calculated per participant and condition, because aggregation with the median is less sensitive to outlier responses than aggregation with the mean (cf. Leys, Ley, Klein, Bernard, & Licata, 2013).

In an additional analysis, we took advantage of the continuous distribution of the response errors to explore whether the LOP effect was based on enhanced access to memory or more precise memory representations in the deep LOP condition. Thus, we fitted the standard mixture model to the data (Brady et al., 2013; Zhang & Luck, 2008) in order to estimate two parameters: the guess rate *g*, reflected by a uniform distribution, and precision *SD*, reflected by the standard deviation of a Gaussian-like distribution. We fitted the data from both LOP conditions simultaneously with the factorial design method provided in the R-package CatContModel (Hardman, 2017). The model fitting procedure implemented a Bayesian Hierarchical framework, using Markov chain Monte Carlo sampling in three chains of 10'000 iterations. The first 500 iterations were removed as burn-in. Non-informative priors were used (cf. Hardman, Vergauwe and Ricker 2017).

We report mean posteriors and 95% credible intervals for the model parameters *g* and *SD*. Comparison of LOP conditions was based on 1000 iterations, the effects are reported as Bayesian Factors (BFs) with 95% Bayesian credible intervals. BF_{10} indicates the likelihood ratio for evidence for the alternative hypothesis over the null hypothesis. According to Jeffreys (Jeffreys, 1961, in Wetzels et al., 2011), BFs from 1 - 3 are regarded as anecdotal, 3 - 10 as moderate, 10 - 30 as strong, 30 -100 as very strong and higher than 100 as extreme evidence in favor of the alternative hypothesis.

Results

Averaged medians of the absolute response errors as a function of LOP condition are depicted in Figure 4a. Three participants were excluded because the distribution of their responses was not significantly different from a uniform distribution in either of the LOP conditions (reflecting random guesses, cf. Haberman et al., 2015). For the remaining 55 participants, a *t*-test for paired samples revealed significant greater response errors in the shallow compared to the deep condition, $t(54) = 3.20$, $p = .002$, $d_z = 0.43$ [0.16, 0.77].[1] This indicates a clear LOP effect for cued recall of the color of concrete objects.

[1] A similar effect was observed, when all 58 participants were included, $t(57) = 3.61$, $p = .001$, $d_z = 0.47$ [0.24, 0.76].

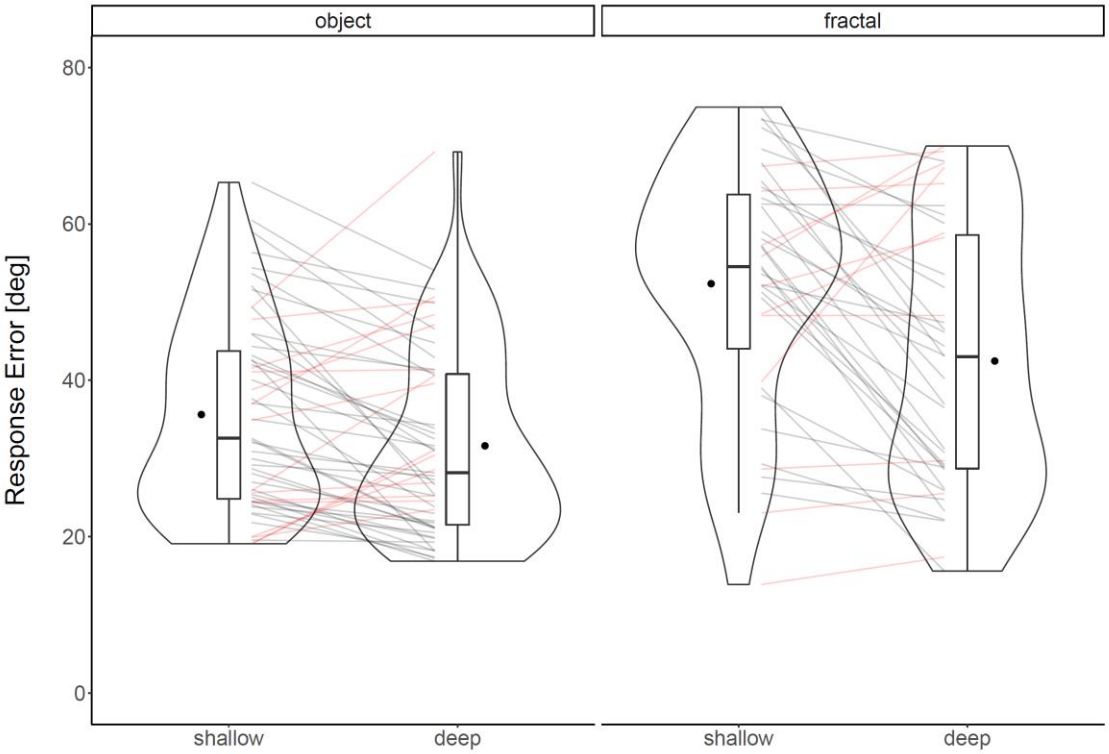

Figure 4. Response error as a function of levels of processing (LOP) condition (shallow, deep), separately for Experiment 3 (objects; left) and Experiment 4 (fractals; right). Lower values represent better performance. The horizontal line inside each box plot indicates median performance (i.e., 50th percentile), and the dot next to each box plot represents average performance. The lower and upper ends of each box represent the 25th and 75th percentiles, enclosing the interquartile range. The whiskers extend to 1.5 times the interquartile range from either end of the box. The violin plots indicate the density of the data. The colored lines indicate change in performance of individual participants between the two LOP conditions. Light gray lines represent participants whose performance improved from the shallow to the deep condition. Light red lines represent participants whose performance decreased from the shallow to the deep condition.

Next, we fitted the standard mixture model to estimate the parameters reflecting guess rate and precision (Hardman et al., 2017; Zhang & Luck, 2008). For parameter *g*, BF_{10} = 5187 [4868, 5520] indicated extreme evidence for a LOP effect with lower guess rate in the deep (*M* = 0.34 [0.29, 0.39]) compared to the shallow (*M* = 0.42 [0.48, 0.37]) condition. For parameter *SD*, no evidence for either hypothesis was indicated, BF_{10} =

0.37 [0.28, 0.48], with a mean posterior of 28.69 degrees [27.18, 30.35] for the deep and a mean posterior of 28.81 degrees [27.16, 30.60] for the shallow condition Thus, LOP seems to affect recall based on access to memory representations but not precision of memory representations.

Experiment 4

Experiment 3 provided strong evidence for an LOP effect on associations between basic visual perceptual features (i.e., color) and concrete objects in a cued recall paradigm with a continuous response space. In Experiment 4 we aimed to replicate and extend these findings with abstract stimuli, because abstract stimuli are hard to verbalize. Experiment 4 will thus again further corroborate whether LOP can indeed affect perceptual features in visual-associative memory. We conducted the same task as in Experiment 3 with abstract fractals instead of concrete objects. Where not otherwise mentioned, methods and analyses were identical to Experiment 3.

Methods

Participants. Due to the observation of close to medium-sized effects in Experiment 3 (Cohen's d_z = .43 and .47, respectively), we based our sample size determination on the a priori power analysis of Experiment 3 and tested another 60 participants. The data of one participant were lost due to a technical error. The remaining 59 participants had a mean age of 23.53 years (*SD* = 3.52, min = 19, max = 34), 33 were women, 48 were right-handed.

Material. Stimuli consisted of 243 abstract patterns, randomly chosen from a pool of 800 fractal patterns. Stimulus preparation followed the same procedure as in Experiment 2. All stimuli fitted into 3.5 x 3.5 degrees of visual angle and were surrounded by a black circle (representing the color wheel) with a radius of 2.5 degrees of visual angle The stimuli were randomly assigned to a practice list of three stimuli and two sets of 120 stimuli each, to be used in the shallow or deep condition, respectively. The stimuli of each set were used to create 24 stimulus lists of five stimuli. Correspondingly, a practice list of three colors and two sets of 120 colors each were created. The colors of both sets were equidistantly sampled on an isoluminant color wheel in CIELUV color space. The colors of the two sets differed by one degree on the color wheel. The colors of each set were further assigned to 24 lists of five colors each. The colors in each list were always as distinct as possible.

Procedure. The procedure was identical to Experiment 3. Because fractals are more difficult to memorize than objects, blocks were reduced to five encoding and recall trials to prevent floor effects. Example encoding and recall trials are depicted in Figures 3c and d. Both LOP conditions (deep and shallow) consisted of 24 blocks each. That is, the number of stimuli per condition was the same as for Experiment 3.

Design and analyses. The design was identical to Experiment 3. We conducted the same analysis as for Experiment 3.

Results

Averaged medians of absolute response errors as a function of LOP condition are depicted in Figure 4c. Twenty participants were excluded because the distribution of their responses was not significantly different from a uniform distribution in either of the LOP conditions (reflecting pure guesses, cf. Haberman et al., 2015). For the remaining 39 participants, a *t*-test for paired samples revealed significant greater response errors in the shallow compared to the deep condition, $t(38) = 4.25$, $p < .001$, $d_z = 0.68$ [0.35, 1.10.

Considering also the slightly lower effect sizes when all participants were included,[2] this indicates a medium to large LOP effect for cued recall of the color of abstract fractals.

Next, we fitted the standard mixture model to estimate the parameters reflecting guess rate and precision (Zhang & Luck, 2008), using the same approach as in Experiment 3. For parameter *g*, the BF_{10} = 69706 [65347, 74142] indicated extreme evidence for a LOP effect in the given dataset, with lower guess rate in the deep (*M* = 0.58 [0.51, 0.64]) than in the shallow condition (*M* = 0.68 [0.62, 0.74]). For parameter *SD*, the BF_{10} = 1.08 [0.83, 1.44], indicated no conclusive evidence for a difference in precision of the remembered items between LOP conditions, with a mean posterior of 23.64 degrees [21.88, 25.71] in the deep and a mean posterior of 24.87 degrees [22.69, 27.28]. Thus, also for abstract fractal-color associations, LOP seems to affect recall based on access to memory representations but not precision of memory representations.

Discussion

Across four experiments, we were able to unambiguously demonstrate for the first time that different levels of processing affect memory for basic visual perceptual features (i.e., color). In Experiment 1 and 2, we showed that deeper LOP benefit recognition memory for concrete object-color associations and abstract fractal-color associations, respectively. In Experiment 3 and 4, we showed that deeper LOP benefit cued recall for concrete object-color associations and abstract fractal-color associations, respectively. Taking advantage of the continuous response space in Experiment 3 and 4, we were further able to show by means of a computational modelling approach that deeper LOP

[2] A similar effect was observed, when all 59 participants were included, $t(58) = 3.91$, $p < .001$, $d_z = 0.51$ [0.27, 0.81].

enhance access to memory representations, but not their mnemonic precision. The results are also in line with subjective reports that the deep encoding condition was easier than the shallow condition.

The effect sizes that we observed across our experiments (Cohen's d_z from 0.43 to 0.78) seem extremely reliable in light of the fact that we assessed memory for basic visual perceptual features. From a critical viewpoint, one might argue that objects and fractals can be verbalized. This is also what the participants reported (Exp. 1 and 2). Nevertheless, memory for abstract fractals (Exp. 2 and 4) was generally lower than memory for concrete objects (Exp. 1 and 3). This suggests that it was harder to verbalize abstract fractals than concrete objects. Interestingly, LOP effects tended to be numerically larger for fractals (Cohen's d_z from 0.68 to 0.78) than objects (Cohen's d_z from 0.41 to 0.74). Similarly, one might argue that color can be verbalized. The German language has eleven basic color terms (the same applies to English; Berlin & Kay, 1969; Witzel & Gegenfurtner, 2018) and there were eight color terms (pink, red, orange, yellow, green, blue, purple, and brown) that could be used to refer to the colors that we sampled (grey, white and black were not sampled). It was not possible to use attributes such as light and dark because we rendered isoluminant colors. Moreover, isoluminant colors do not have clear category boundaries (e.g., between green and blue), rendering it more difficult to attribute unambiguous labels to colors. We cannot entirely rule out the possibility that verbal strategies were applied to memorize object-color associations. However, verbal strategies are unlikely to be successful for fractal-color associations, due to the low distinctiveness of the fractals. An impressive demonstration of this explanation can be observed in the color lure condition with fractals. Performance was at chance level in the shallow condition (where participants had to indicate the presence of straight lines during encoding) and increased above chance level in the deep condition (where participants had to rate the pleasantness of the stimuli during encoding). LOP effects, as observed across our tasks, were thus at least partly depending on visual rather than verbal representations.

Computational modelling the results of Experiments 3 and 4 revealed extreme evidence that deeper levels of processing enhance access to visual memory representations but revealed no clear effect on mnemonic precision. One could reasonably argue that the shallow encoding instruction in our work (i.e., to indicate if a stimulus consists of a straight line) resulted in a narrow attentional focus on the details of a stimulus. By contrast, the deep encoding instruction (i.e., to indicate if a stimulus is pleasant or not) resulted in a more global attentional focus. This is in line with previous explanations for LOP effects with faces: Bower and Karlin (1974) found that deep encoding questions (e.g., honesty or likability judgement) led to better memory for face stimuli compared to shallow encoding questions (e.g., gender judgement). They suggested that honesty and likability judgements, due to the vague judgement criteria, lead to a holistic focus and thus the extraction of more details. By contrast, gender judgements, due to the clear judgement criteria, can be based on one or the other salient cue and thus lead to a narrow focus and the extraction of less details. Notably, these explanations are closely linked to the verbal overshadowing effect, where giving a verbal description of a stimulus (i.e., accessing its salient features) impairs later recognition (Schooler & Engstler-Schooler, 1990). However, previous work manipulating the emotional context of stimuli, found that emotional vs. neutral context increased mnemonic precision, due to a narrower attentional focus in emotionally valent contexts

(Xie & Zhang, 2017b, 2017a). Thus, it will be an interesting endeavor for future research to disentangle how LOP manipulations with images might affect attentional focus and how this impacts access to memory or mnemonic precision.

Our work has also practical implications, given the importance of memory based on visual impressions in everyday life. For instance, individuals tend to focus on details of a visual scene when trying to memorize it very well. While it seems very reasonable to focus on the details when attempting to retain detailed memories of an experience, our work indicates that this is unlikely to be the best strategy. By contrast, our findings unequivocally suggest that it is more beneficial to ask if you like what you are seeing. This may seem counter-intuitive but finds further indirect support in studies showing that personally relevant material is better remembered than neutral material (e.g., Levine & Edelstein, 2009). The question whether a stimulus is appealing might render it personally more relevant in contrast to an analytic question with regards to the presence of some basic features in a stimulus and will thus further support deeper levels of processing.

To summarize and conclude, we demonstrated extremely reliable LOP effects for basic perceptual features in visual-associative memory across a series of four statistically well-powered experiments. Computational modelling revealed that the effects were due to enhanced access to visual memory rather than enhanced fidelity of visual memory representations. The findings indicate that the hitherto influential LOP account is much more universal than the focus on verbal stimuli in the current literature might suggest.

References

Baddeley, A. D., & Hitch, G. J. (2017). Is the Levels of Processing effect language-limited? *Journal of Memory and Language*, *92*, 1–13. https://doi.org/10.1016/j.jml.2016.05.001

Berlin, B., & Kay, P. (1969). *Basic color terms: their universality and evolution*. Berkley, CA: University of California Press.

Bower, G. H., & Karlin, M. B. (1974). Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology*, *103*(4), 751–757. https://doi.org/10.1037/h0037190

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

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The bank of standardized stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS ONE*, *5*(5), e10773. https://doi.org/10.1371/journal.pone.0010773

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684.

D'Agostino, P. R., O'Neill, B. J., & Paivio, A. (1977). Memory for pictures and words as a function of level of processing: depth or dual coding? *Memory & Cognition*, *5*(2), 252–256. https://doi.org/10.3758/BF03197370

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Haberman, J., Brady, T. F., & Alvarez, G. A. (2015). Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *Journal of Experimental Psychology: General*, *144*(2), 432–446. https://doi.org/10.1037/xge0000053

Hardman, K. O. (2017). CatContModel: Categorical and Continuous Working Memory Models for Delayed-Estimation Tasks.

Hardman, K. O., Vergauwe, E., & Ricker, T. J. (2017). Categorical Working Memory Representations are used in Delayed Estimation of Continuous Colors. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(1), 30–54. https://doi.org/http://dx.doi.org/10.1037/xhp0000290

Intraub, H., & Nicklos, S. (1985). Levels of processing and picture memory: the physical superiority effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(2), 284–298.

Jeffreys, H. (1961). *Theory of probability*. Oxford: Oxford University Press.

Kassambara, A. (2019). rstatix: Pipe-Friendly Framework for Basic Statistical Tests. Retrieved from https://cran.r-project.org/package=rstatix

Lakens, D., & Caldwell, A. R. (2019). Simulation-based power-analysis for factorial ANOVA designs. *PsyArXiv*, 1–8. https://doi.org/10.31234/osf.io/baxsff

Levine, L. J., & Edelstein, R. S. (2009). Emotion and memory narrowing: A review and goal-relevance approach. *Cognition and Emotion*, *23*(5), 833–875. https://doi.org/10.1080/02699930902738863

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Ovalle Fresa, R., & Rothen, N. (2019). Training enhances fidelity of color representations in visual long-term memory. *Journal of Cognitive Enhancement*, *3*(3), 315–327. https://doi.org/10.1007/s41465-019-00121-y

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/10.1163/156856897X00366

R Core Team, R. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: some things are better left unsaid. *Cognitive Psychology*, *22*, 36–71. https://doi.org/10.1016/0010-0285(90)90003-M

Sutterer, D. W., & Awh, E. (2016). Retrieval practice enhances the accessibility but not the quality of memory. *Psychonomic Bulletin and Review*, *23*(3), 831–841. https://doi.org/10.3758/s13423-015-0937-x

Wesemann, W., Schiefer, U., & Bach, M. (2010). Neue DIN-Normen zur Sehschärfebestimmung. *Ophthalmologe*, *107*(9), 821–826.

https://doi.org/10.1007/s00347-010-2228-2

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science*, *6*(3), 291–298. https://doi.org/10.1177/1745691611406923

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: the SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684. https://doi.org/10.3758/BRM.42.3.671

Witzel, C., & Gegenfurtner, K. R. (2018). Color Perception: Objects, Constancy, and Categories. *Annual Review of Vision Science*, *4*(1), 475–499. https://doi.org/10.1146/annurev-vision-091517-034231

Xie, W., & Zhang, W. (2017a). Mood-dependent retrieval in visual long-term memory: dissociable effects on retrieval probability and mnemonic precision. *Cognition and Emotion*, *32*(4), 674–690. https://doi.org/10.1080/02699931.2017.1340261

Xie, W., & Zhang, W. (2017b). Negative emotion enhances mnemonic precision and subjective feelings of remembering in visual long-term memory. *Cognition*, *166*, 73–83. https://doi.org/10.1016/j.cognition.2017.05.025

Xu, X., Zhao, Y., Zhao, P., & Yang, J. (2011). Effects of level of processing on emotional memory: Gist and details. *Cognition and Emotion*, *25*(1), 53–72. https://doi.org/10.1080/02699931003633805

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233–235. https://doi.org/10.1038/Nature06860