



Pictorial low-level features in mental images: evidence from eye fixations

Corinna S. Martarelli¹ · Fred W. Mast²

Received: 7 September 2020 / Accepted: 26 February 2021 / Published online: 22 March 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

It is known that eye movements during object imagery reflect areas visited during encoding. But will eye movements also reflect pictorial low-level features of imagined stimuli? In this paper, three experiments are reported in which we investigate whether low-level properties of mental images elicit specific eye movements. Based on the conceptualization of mental images as depictive representations, we expected low-level visual features to influence eye fixations during mental imagery, in the absence of a visual input. In a first experiment, twenty-five participants performed a visual imagery task with high vs. low spatial frequency and high vs. low contrast gratings. We found that both during visual perception and during mental imagery, first fixations were more often allocated to the low spatial frequency–high contrast grating, thus showing that eye fixations were influenced not only by physical properties of visual stimuli but also by its imagined counterpart. In a second experiment, twenty-two participants imagined high contrast and low contrast stimuli that they had not encoded before. Again, participants allocated more fixations to the high contrast mental images than to the low contrast mental images. In a third experiment, we ruled out task difficulty as confounding variable. Our results reveal that low-level visual features are represented in the mind's eye and thus, they contribute to the characterization of mental images in terms of how much perceptual information is re-instantiated during mental imagery.

Introduction

Visual perception is influenced by the sensory input and cognitive top-down processes (Albright, 2012); whereas, mental imagery does not require a sensory input but still draws on some of the same mechanisms that are also involved in the process of visual perception. A case in point are eye movements. Early research in vision science (Yarbus, 1967) has shown that during visual perception, eye movements are influenced by stimulus properties and by the task at hand. Subsequent studies emphasizing bottom-up attention guidance showed that eye fixations are partly triggered by pictorial low-level features, such as local contrast, edges, or a sudden movement (Borji et al., 2013; Henderson et al., 1999; Itti, 2005; Kienzle et al., 2009; Le Meur et al., 2007). Various suggestions have been made on the type of low-level

features that mostly attract fixations. For example, a saliency map predicts fixated locations based on low-level features (Itti & Koch, 2000, 2001; Koch & Ullman, 1985; Parkhurst et al., 2002). These computational models ascribed a predominant role to contrast and edges (e.g., Itti & Koch, 2000). It has to be noted, however, that Mannan et al. (1996) raised doubt that the pattern of fixations during image viewing can be associated with locations of low-level features such as contrast, luminance, spatial frequency or edge density [see also Einhäuser and König (2003) for a similar finding]. More recent work using a Bayesian approach that accounts for correlations between features has shown that high frequency edges attract fixations (Baddeley & Tatler, 2006). Since the early work by Koch and Ullman (1985), many computational models of bottom-up attention have emerged. These models put forward low-level features as fundamental building blocks of visual perception. For a systematic review, see for example Itti and Borji (2014).

Interestingly, the impact of pictorial low-level features on eye fixations during mental imagery is not known, despite the fact that mental images have picture-like qualities presumably containing low-level features (e.g., Kosslyn et al., 2006; Rouw et al., 1997; see also more recent studies using

✉ Corinna S. Martarelli
corinna.martarelli@fernuni.ch

¹ Faculty of Psychology, Swiss Distance University Institute, Brig, Switzerland

² Department of Psychology, University of Bern, Bern, Switzerland

multivariate pattern analysis, e.g., Albers et al., 2013; Johnson & Johnson, 2014; Lee et al., 2012; Naselaris et al., 2015). Evidently, there is no low-level bottom-up input that can trigger eye movements during mental imagery. Nevertheless, certain features in mental images attract more fixations than others. It is known that eye movements are not only involved in the processing of visual information but also during mental imagery without obvious reason to move one's eyes. Several studies demonstrate that empty areas visited during imagery and memory tasks reflect locations that were inspected during prior encoding of the stimuli (Altmann, 2004; Bone et al., 2019; Brandt & Stark, 1997; Chiquet et al., 2020; Fourtassi et al., 2013; Johansson et al., 2006; Johansson et al., 2012; Johansson & Johansson, 2014; Laeng & Teodorescu, 2002; Laeng et al., 2014; Martarelli & Mast, 2011, 2013; Martarelli et al., 2017; Richardson & Spivey, 2000; Scholz et al., 2016; Scholz et al., 2018; Spivey & Geng, 2001; Umar et al., 2021). However, most of these studies asked participants to visualize previously seen objects presented on a blank screen, without considering the low-level feature content of mental objects. An issue is that different low-level features tend to be correlated where the objects are presented; for example, low-level features often change at the border of objects (Einhäuser et al., 2008; Schütz et al., 2011). Here, we sought to distinguish between eye fixations to mental images of objects and eye fixations to imagined low-level features. We aimed to characterize mental images in terms of how much low-level information is re-instantiated during mental imagery.

In Experiment 1, participants were asked to imagine an array of gratings which differed in spatial frequency and contrast (high vs. low, similar to the study by Kosslyn et al., 1999). We used eye movements as an indicator of saliency both during perception and during mental imagery. We use the original meaning of the term saliency by referring to locations that stick out from the display and attract attention (Koch & Ullman, 1985). We expected eye movements to be influenced not only by physical properties of the stimuli but also by their imagined counterpart. Eye fixations during picture viewing are partly triggered by low-level features, and the aim of this study is to investigate whether eye movements accompany low-level properties when the same information is not seen but imagined.

Experiment 1

Method

Participants

An a priori $G \times$ Power analysis revealed that we needed a sample of 24 participants to detect a medium to large

effect (2×2 analysis of variance, within-subjects factors; parameters: partial $\eta^2 = 0.13$, $\alpha = 0.05$, $1 - \beta = 0.95$; Faul et al., 2007). Twenty-five adults (17 female, aged between 20 and 31 years) participated in the experiment. All participants were naïve about the purpose of the experiment and had normal or corrected-to-normal visual acuity. The study (including Experiments 1–3) was approved by the university's local ethics committee and was carried out in accordance with the Declaration of Helsinki of 1975. The datasets (Experiments 1–3) are available from Figshare (<https://figshare.com/s/e64c369ef8a5dbb2ed27>).

Apparatus

Eye movements were recorded using an SMI RED tracking system (SensoMotoric Instruments, Teltow, Germany). Data were registered with a sampling rate of 50 Hz, a spatial resolution of 0.1° and a gaze position accuracy of 0.5° . The eye-tracking device is contact free and determines the gaze on the image by combining the cornea reflex with the pupil location, via an infrared light sensitive video camera. The stimuli were presented on a 17-in. screen using Experiment Center Software and eye data were recorded with I-View X Software, both developed by SensoMotoric Instruments. Participants went through a five-point calibration procedure, followed by a four-point validation. The average horizontal and vertical error was smaller than 0.8° angular accuracy.

Materials

Participants were seated at approximately 50 cm from the screen. The screen was divided by the vertical and horizontal midlines into four equally sized areas of interest (AOIs). Numbers could appear in each area (number 1: upper left area; number 2: upper right area; number 3: lower left area; number 4: lower right area). Grayscale sinusoidal gratings were generated using the Online-Software Gabor Patch Generator (www.cogsci.nl/software/online-gabor-patch-generator). The gratings were presented in a circular aperture (visual angle 10°) with high (RGB = 255/255/255) vs. low contrast (RGB = 128/128/128), and high (0.33 cycles/ $^\circ$) vs. low spatial frequency (0.20 cycles/ $^\circ$). Each array contained the four gratings (two contrast levels and two spatial frequencies) in all possible position-combinations (24 arrays). We also varied the orientation of the gratings (45° vs. 135°). Each array contained only one orientation. In total, there were forty-eight arrays of the four gratings (24 possible position-combinations of the four gratings in two different orientations, 45° and 135°). See Fig. 1a (original image) for an example of the stimuli.

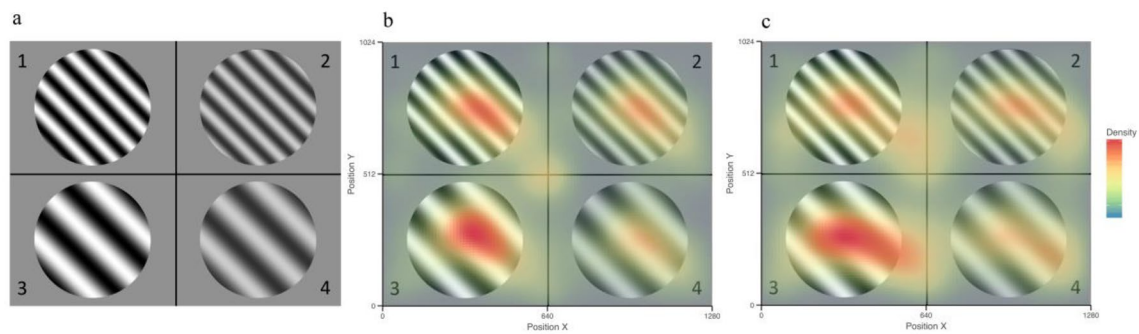


Fig. 1 **a** Original image; **b** Heat map of the first three fixations during perception across all trials and all participants; **c** Heat map of the first three fixations during imagery across all trials and all participants

Procedure

The procedure was adapted from Kosslyn et al. (1999). Participants were given instructions on the computer prior to the task. They learned the configuration to remember which number labeled each quadrant. The perceptual encoding phase was composed of forty-eight different arrays of gratings, each of which was presented for 6 s. After each presentation of an array, participants had to imagine the gratings they just saw (image generation) and then answer to a specific question (image inspection). In the image generation phase, only the cross was present on a gray background and participants heard a sound “Imagine” (all sound files were created by means of Audacity, <http://audacity.sourceforge.net>), which cued them to create a vivid image of the previously seen array of gratings. As soon as they generated the mental image of the entire array of stimuli, they pressed the space bar to begin with the image inspection phase. Participants listened to prerecorded voices indicating a sequence of two numbers followed by the name of one of two properties, such as “1, 2, contrast” or “1, 2, space”. The numbers indicated the two areas that were to be compared, and the property indicated the comparison the participant had to make. We used two times the 24 sound files (all the possible combinations), and randomly associated them to the arrays of gratings within the two orientations (45° and 135°). Participants had to answer by operating one of two buttons (“c” and “m”) on the keyboard.

The image generation phase lasted 5.3 s on average ($SD = 1794$ ms; mean number of fixations = 13.4, $SD = 5.3$). We focused on the comparison of fixations during the image generation phase and during visual perception. The image inspection phase was primarily carried out to assure that participants were on task. It also reduced the possibility that participants figured out the purpose of the experiment and behaved accordingly. At the end of the testing session, participants were asked to guess the hypothesis of the study.

Results

Fixations were detected when the sum of the dispersion of the gaze stream on the x and y axes was below 100 pixels and when the duration exceeded 80 ms. Fixation times below 100 ms were removed from the analyses. The four quadrants were defined as AOIs and we compared the first fixation, number of fixations, and time spent in the four quadrants (i.e., high spatial frequency–high contrast grating, low spatial frequency–high contrast grating, high spatial frequency–low contrast grating, and low spatial frequency–low contrast grating). Note that since the gratings were presented in each area 12 times, any absolute looking bias is distributed equally across the four gratings.

Based on computational models of bottom-up attention (e.g., Parkhurst et al., 2002), we were mainly interested in the position of the first fixations (see Fig. 1). However, a visual exploration of all fixations (see Appendix A) illustrated that eye behavior was very similar independent of the order of fixations. Thus, we decided to carry out analyses not only on first fixations but also overall on number of fixations and time spent during perception and during imagery.

Participants were able to solve the task (image inspection phase) and they were correct in 70% ($SD = 21\%$) of the trials. The level of accuracy was comparable to the findings of Thompson et al. (2001) who used a similar task (in their study mean error rates were 29.1% for the high-resolution arrays and 24.9% for the low-resolution arrays). We included all data of the image generation phase in the analyses reported below. Excluding participants performing at chance level (5 out of 25) or excluding trials of the image generation phase that were solved incorrectly in the image inspection phase did not yield substantial differences in the analysis of eye movements. Moreover, no participant correctly guessed the hypothesis of the study. Participants principally referred to the image inspection phase and thought that the hypothesis was about performance in this task. The values of asymmetry and of kurtosis for the different variables were

between -1.30 and 3.18, thus considered acceptable to prove normal univariate distribution. The analyses were computed with IBM SPSS version 25. We computed analyses of variance and report partial eta squared as measures of effect size.

First fixations during perception and imagery

A 2 × 2 analysis of variance with the within-subject factors spatial frequency (low/high) and contrast (low/high), and the percentage of the first fixations in the four AOIs as dependent variable revealed that participants allocated more first fixations to the *low spatial frequency–high contrast grating* (28.03% during perception and 31.59% during imagery; chance level = 25%) compared to the other three gratings (means of percentages are reported in Table 1). Indeed, the interactions turned out to be significant (during perception: $F(1, 24) = 4.48, p = 0.045, \text{partial } \eta^2 = 0.16$; during imagery: $F(1, 24) = 6.58, p = 0.017, \text{partial } \eta^2 = 0.21$). During perception, the main effect contrast also turned out to be significant ($F(1, 24) = 8.38, p = 0.008, \text{partial } \eta^2 = 0.26$), whereas the main effect spatial frequency was not significant ($F < 1$). During imagery, both main effects were not significant ($ps > 0.075$).

We evaluated the distribution of the first three fixations. The middle panel of Fig. 1 shows the cumulative spatial distribution of the first three fixations during visual perception across all trials and all participants and the right panel of Fig. 1 illustrates the cumulative spatial distribution of the first three fixations during mental imagery (blank screen) across all trials and all participants. Red regions were fixated more frequently. Visual inspection of the heat maps suggests

that participants allocated more first fixations to the *low spatial frequency–high contrast grating* both during perception and during imagery.

Further, we considered the last fixation during the preceding perception phase. Numerically, the same pattern of results was revealed (high spatial frequency–high contrast grating = 24.1%, $SD = 9.3\%$, high spatial frequency–low contrast grating = 23.5%, $SD = 9.8\%$, low spatial frequency–high contrast grating = 27.7%, $SD = 9.1\%$, low spatial frequency–low contrast grating = 24.8%, $SD = 7.6\%$); however, the results turned out to be non-significant, when analyzed with a 2 × 2 analysis of variance with the within-subject factors spatial frequency (low/high) and contrast (low/high), and the percentage of the last fixations in the four AOIs in the end of the perception phase as dependent variable ($ps > 0.153$).

Number of fixations and time spent in the four AOIs during perception and imagery

We conducted analyses of variance with spatial frequency (low/high) and contrast (low/high) as within-subject factors and percentages of number of fixations and percentage of time spent in the four AOIs as dependent variables (in separate models, one dependent variable for each model), both during perception and during imagery. Means and standard error of the mean (percentages) are reported in Table 1. Computing the same analyses with absolute values (number of fixations and time spent) did not change the findings in terms of accepting the null hypothesis. Absolute values are reported in Appendix B.

Table 1 Percentages of first fixations, fixation time, and number of fixations for high/low spatial frequency and high/low contrast gratings separated by task (perception and imagery)

Dependent variable	Grating			
	highSF highCON	highSF lowCON	lowSF highCON	lowSF lowCON
Percentage of first fixations (perception)	$M = 25.54 (SEM = 0.98)$	$M = 24.68 (SEM = 0.96)$	$M = 28.03^{**} (SEM = 1.02)$	$M = 21.75^{**} (SEM = 0.95)$
Percentage of number of fixations (perception)	$M = 25.68 (SEM = 1.06)$	$M = 21.29^{**} (SEM = 1.04)$	$M = 29.75^{***} (SEM = 0.95)$	$M = 23.27 (SEM = 0.93)$
Percentage of time spent (perception)	$M = 25.47 (SEM = 0.97)$	$M = 21.44^{**} (SEM = 0.96)$	$M = 29.69^{***} (SEM = 0.96)$	$M = 23.39 (SEM = 0.87)$
Percentage of first fixations (imagery)	$M = 22.24 (SEM = 2.29)$	$M = 23.32 (SEM = 2.22)$	$M = 31.59^{**} (SEM = 1.78)$	$M = 22.84 (SEM = 1.76)$
Percentage of number of fixations (imagery)	$M = 25.44 (SEM = 1.35)$	$M = 22.37 (SEM = 1.47)$	$M = 28.99^{**} (SEM = 1.40)$	$M = 23.19 (SEM = 1.15)$
Percentage of time spent (imagery)	$M = 23.95 (SEM = 1.36)$	$M = 23.47 (SEM = 1.56)$	$M = 29.72^{**} (SEM = 1.61)$	$M = 22.86 (SEM = 1.16)$

highSF highCON spatial frequency high–contrast high, *highSF lowCON* spatial frequency high–contrast low, *lowSF highCON* spatial frequency high–contrast high, *lowSF lowCON* spatial frequency low–contrast low, M = mean, SEM = standard error of the mean

Percentages that differ from chance level (25%) are marked with ** when $p < 0.01$, and *** when $p < 0.001$ (uncorrected one-sample t -tests).

During perception, participants allocated more fixations to the *low spatial frequency–high contrast grating* compared to the other three gratings (means are reported in Table 1) as revealed by a significant interaction ($F(1, 24) = 4.42$, $p = 0.046$, partial $\eta^2 = 0.16$). The main effect of contrast ($F(1, 24) = 16.89$, $p < 0.001$, partial $\eta^2 = 0.41$) turned out to be significant as well as the main effect of spatial frequency ($F(1, 24) = 6.18$, $p = 0.020$, partial $\eta^2 = 0.20$). The main effects showed that participants allocated significantly more fixations to the high contrast gratings ($M = 27.58$, $SEM = 0.63$) than to the low contrast gratings ($M = 22.42$, $SEM = 0.63$), and significantly more fixations to the low spatial frequency gratings ($M = 26.54$, $SEM = 0.62$) than to the high spatial frequency gratings ($M = 23.46$, $SEM = 0.62$). The analysis of time spent during perception revealed a main effect of contrast ($F(1, 24) = 15.66$, $p = 0.001$, partial $\eta^2 = 0.39$) and a main effect of spatial frequency ($F(1, 24) = 5.24$, $p = 0.031$, partial $\eta^2 = 0.19$), whereas the interaction turned out to be non-significant ($F(1, 24) = 3.20$, $p = 0.086$, partial $\eta^2 = 0.12$). The main effects showed a similar pattern as the analysis of number of fixations. Participants spent more time in the high contrast grating ($M = 27.72$, $SEM = 0.69$) than in the low contrast grating ($M = 22.28$, $SEM = 0.62$) and they spent significantly more time in the low spatial frequency grating ($M = 26.51$, $SEM = 0.66$) than in the high spatial frequency grating ($M = 23.49$, $SEM = 0.66$).

During mental imagery, the analysis of number of fixations revealed a significant effect of contrast ($F(1, 24) = 6.21$, $p = 0.020$, partial $\eta^2 = 0.21$), whereas the other effects turned out to be non-significant (main effect of spatial frequency: $F(1, 24) = 1.56$, $p = 0.223$, partial $\eta^2 = 0.06$; interaction: $F(1, 24) = 1.75$, $p = 0.199$, partial $\eta^2 = 0.07$). Participants allocated more fixations to the high contrast grating ($M = 27.22$, $SEM = 0.89$) than to the low contrast grating ($M = 22.78$, $SEM = 0.89$).

Participants spent more time in the imagined *low spatial frequency – high contrast gratings* compared to the other three gratings as revealed by a significant interaction ($F(1, 24) = 8.50$, $p = 0.008$, partial $\eta^2 = 0.26$), means are reported in Table 1. The main effects turned out to be non-significant (contrast: $F(1, 24) = 4.17$, $p = 0.052$, partial $\eta^2 = 0.15$; spatial frequency: $F(1, 24) = 1.76$, $p = 0.196$, partial $\eta^2 = 0.07$).

Discussion

The results from Experiment 1 show that low-level features elicited specific eye movements not only during the process of perception but also during mental imagery. Participants allocated the first fixations more often to the *low spatial frequency–high contrast grating* both during visual perception and during mental imagery. Analyses of time spent, and number of fixations show a similar result. The findings support the account of imagery as perception-like; pictorial

low-level features are preserved in mental imagery (Broggin et al., 2012; Laeng & Sulutvedt, 2014; Rouw et al., 1997). As participants did not guess the hypothesis, it seemed unlikely that they more frequently looked at the *low spatial frequency–high contrast grating* to comply with the experiment's purpose or to behave as if they were in the corresponding perceptual situation.

Previous studies revealed that eye movements during mental imagery (number of fixations and time spent) revisit the locations where the stimuli were encoded (Altmann, 2004; Brandt & Stark, 1997; Fourtassi et al., 2013; Johansson & Johansson, 2014; Johansson et al., 2006, 2012; Laeng & Teodorescu, 2002; Laeng et al., 2014; Martarelli & Mast, 2011, 2013; Martarelli et al., 2017; Richardson & Spivey, 2000; Scholz et al., 2016, 2018; Spivey & Geng, 2001). It has been concluded that the spatial location is integrated in the memory trace along visual and semantic information. This research is important because it provides a possible interpretation of the eye pattern that we found in Experiment 1. During mental image generation, participants could look back to the areas visited during perceptual encoding. In other words, it is conceivable that participants allocated more fixations to the *low spatial frequency–high contrast grating* area during visual mental imagery, because they allocated more fixations to this grating during visual perception and not necessarily because low-level information is preserved in mental images.

It is interesting that the percentages of last fixations during the preceding perception phase were distributed equally among the four AOIs (in terms of significance), thus suggesting that the allocation of first fixations during the subsequent image generation phase was not caused by a residual afterimage. Nevertheless, the occurrence of afterimages during the image generation phase cannot be excluded. Hence, we conducted a second experiment avoiding these potential confounds.

Experiment 2

A possible interpretation of the systematic eye pattern found during mental image generation (Exp. 1) is that during imagery, participants could look back to the areas visited during perceptual encoding. In Experiment 2, we aimed to rule out the potential influence of eye position during encoding. We assessed whether low-level features influence mental imagery by avoiding the influence of short-term memory processes. In Experiment 2, participants were asked to imagine high contrast vs. low contrast images that they had not seen before. Instead of using gratings as we did in the first experiment, we decided to use objects as stimulus material. We needed a larger number of stimuli, and also stimuli that can easily be imagined when not previously seen. Moreover,

using objects might also be more ecological valid, given that in everyday life we rather do not imagine gratings but objects. Given that our participants had to imagine high contrast and low contrast versions of the *same* objects, we do not lose experimental control, which is a typical inherent advantage of using gratings in experimental research.

Method

Participants

Based on the effect sizes of the previous experiment (mean difference between *low spatial frequency–high contrast grating* and *low spatial frequency–low contrast grating*) we computed an a priori $G \times$ Power analysis. The analysis revealed that we needed a sample of at least 18 participants (paired-samples t -test, two-tailed; parameters: $d_z = 0.91$, $\alpha = 0.05$, $1 - \beta = 0.95$; Faul et al., 2007). Twenty-two adults (19 females, aged between 19 and 40 years) participated in the experiment. All participants were naïve about the

purpose of the experiment and had normal or corrected-to-normal visual acuity.

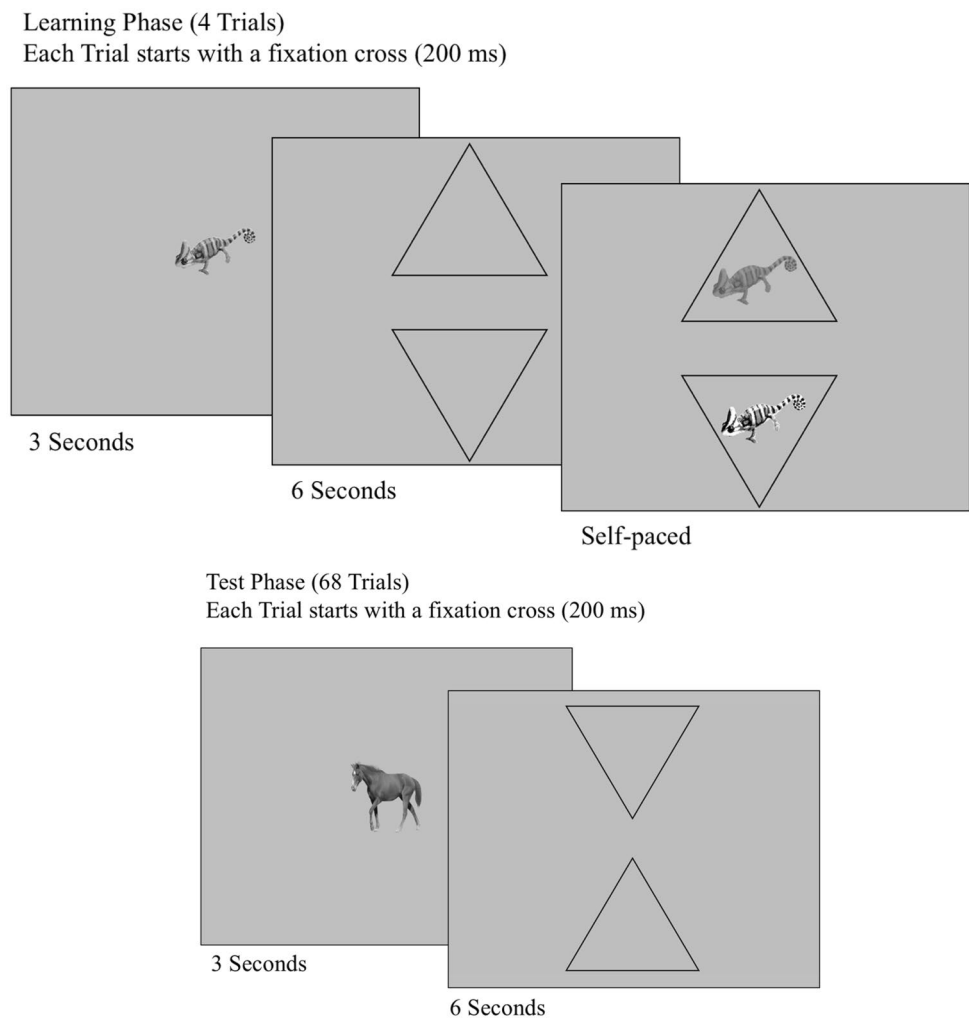
Apparatus

Eye movements were recorded using the same SMI RED tracking system as in Experiment 1.

Materials

The stimuli consisted of 68 images representing animals like an elephant or a buffalo (34 images) or objects like a car or an umbrella (34 images). Low-level stimulus attributes were controlled via the SHINE (spectrum, histogram, and intensity normalization and equalization) toolbox for Matlab developed by Willenbockel et al. (2010). We matched luminance and, thus, minimized potential low-level confounds. An example of images used in the study can be found in Fig. 2. The items were presented in the center of the computer screen against a gray background.

Fig. 2 Illustration of events within one trial in the learning phase (upper part of the figure) and of events within one trial in the study phase (lower part of the figure) in Experiment 2



During the visualization task, participants were shown two triangles, one pointing upward and the other pointing downward. The position of the triangles was counterbalanced, and we had four possible combinations of triangles pointing upward and downward (i.e., (1) triangle pointing upward in the upper-central part of the screen and triangle pointing downward in the lower-central part of the screen; (2) triangle pointing upward in the lower-central part of the screen and triangle pointing downward in the upper-central part of the screen; (3) triangle pointing upward in the left-central part of the screen and triangle pointing downward in the right-central part of the screen; (4) triangle pointing upward in the right-central part of the screen and triangle pointing downward in the left-central part of the screen; see Fig. 2 for two examples). There was no perception condition since we wanted to avoid a memory-based strategy.

Procedure

Participants were seated in front of the computer screen. The distance between participants and the screen was approximately 50 cm. After a 5-point calibration and validation procedure (only error values below 0.8° were accepted), the following task instructions appeared on the screen: “You will be presented with different images (animals or objects). Your task is to visualize the images with high and low contrast. First, you will always see an image with medium contrast. Next, you will see an image with two empty triangles, in which you are going to visualize the image two times simultaneously, a low-contrast version and a high-contrast version of the previously seen image. The triangles indicate where you have to visualize the image with low and high contrast. If the triangle points upward, this means low contrast (*or high contrast*). If the triangle points downward, this means high contrast (*or low contrast*)”.

Learning phase Participants had to learn the association between the triangles (upward vs. downward pointing) and contrast (high vs. low contrast). Half of the participants were told that the upward pointing triangle would indicate high contrast, and a downward pointing triangle low contrast, and vice versa for the other half of participants. In the learning phase, participants were presented with 4 stimuli. They were instructed to visualize the presented objects in high and low contrast and that they would receive feedback after in order to understand what is meant by high and low contrast. They were also instructed that a fixation cross will appear between trials and that they should always fixate the fixation cross. During imagery, they were presented with the four possible displays. After the image generation phase (high contrast and low contrast version of the stimuli), they received feedback and a high contrast (standard deviation of gray levels distribution of

approximately 75) and low contrast (standard deviation of gray levels distribution of approximately 17) image of the original image (standard deviation of gray levels distribution of approximately 48) was presented in the triangles. See an illustration of the learning phase in Fig. 2.

Test phase In the test phase, 68 new images were presented in random order (see Fig. 2). After a fixation cross (200 ms), the image was presented for 3 s; next, one out of four imagery displays appeared (6 s). The four imagery displays (see the Materials section above for a description of the four imagery displays) were presented 17 times each.

Results

As in Experiment 1, analyses were based on time spent, number of fixations, and first fixations.

Time spent and number of fixations during imagery

We computed the mean percentage of time spent on the high contrast triangle and on the low contrast triangle during imagery. The paired-samples *t*-test revealed that participants spent statistically significantly more time in the high contrast triangle ($M = 45.16\%$, $SD = 9.59\%$) than in the low contrast triangle ($M = 38.25\%$, $SD = 9.81\%$), $t(21) = 4.69$, $p < 0.001$, $d = 1.00$.

The same analysis with the percentage of number of fixations allocated to the two triangles showed a similar result. The paired-samples *t*-test revealed that participants allocated statistically significantly more fixations on the high contrast triangle ($M = 40.90\%$, $SD = 8.08\%$) compared to the low contrast triangle ($M = 37.18\%$, $SD = 9.39\%$), $t(21) = 2.50$, $p = 0.021$, $d = 0.53$.

First fixations during imagery

Participants allocated a mean of 9.99 ($SD = 1.98$) fixations per trial (6 s imagery). Figure 3 shows the time spent on the high contrast triangle and on the low contrast triangle for the first 12 fixations (*x*-axis). It is interesting to see that the probability of going to one or the other triangle during the first two fixations was at chance. Starting from the third fixation (until the seventh fixation), participants spent more time in the high contrast triangle than in the low contrast triangle. This pattern seemed to reverse during the last fixations but there were no significant differences. We computed paired-*t* tests (uncorrected) and report them in Table 2. Note that the significant differences reported in Table 2 do not survive a Bonferroni correction

Fig. 3 Mean percentage of time spent for high contrast and low contrast visual mental images. Order of fixation is represented in the x-axis (1 = first fixation until 12 = 12th fixation). Errors bars indicate SEM

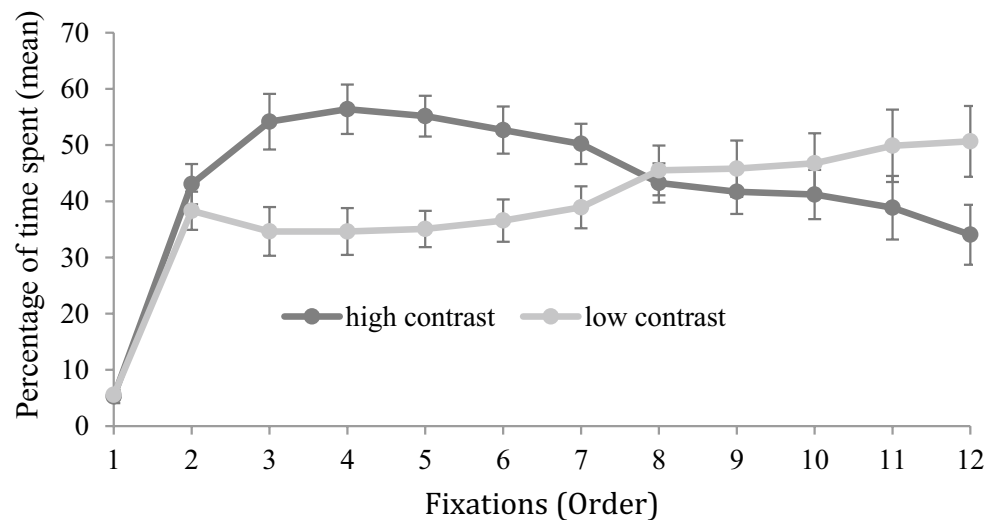


Table 2 Significant pairwise differences are shown in cursive font (uncorrected p-values)

Paired samples <i>T</i> -test	<i>T</i>	<i>df</i>	<i>p</i>	Cohen's <i>d</i>
First fixation	0.37	21	0.718	0.08
Second fixation	0.85	21	0.406	0.18
Third fixation	2.36	21	0.028	0.50
Fourth fixation	2.85	21	0.010	0.61
Fifth fixation	3.63	21	0.002	0.77
Sixth fixation	2.43	21	0.024	0.52
Seventh fixation	1.83	21	0.081	0.39
Eighth fixation	0.33	21	0.746	0.07
Ninth fixation	0.52	21	0.607	0.11
Tenth fixation	0.62	21	0.540	0.13
Eleventh fixation	0.95	21	0.351	0.20
Twelfth fixation	1.56	21	0.133	0.33

(0.05/12 = 0.004). The significant results cluster around the third and six fixations, such that a Bonferroni correction can be considered overly conservative.

Further, Fig. 3 disregards the exact latencies of the fixations. In Appendix C, a plot showing the mean proportion of fixations on imagined high vs. low contrast images per successive 100-ms time bins is reported. This approach suggests an overall predominance of the high contrast images.

Discussion

The results of Experiment 2 confirm the results from Experiment 1; low-level visual features elicit specific eye movements during mental imagery. In Experiment 2, eye position during encoding and afterimages could not influence

eye position during imagery because participants never saw the low and high contrast versions of the stimuli. Therefore, we had no perception condition in Experiment 2. Moreover, the content of the stimuli could not influence eye behavior, because participants had to imagine low and high contrast versions of exactly the same stimuli. Still, participants allocated more fixations to the high contrast mental images than to the low contrast mental images. What is more, fixations in the first part of the trial were allocated more often to high contrast mental images than to low contrast mental images. It has to be noted, however, that the advantage of high contrast was observed only from the third fixation (see Fig. 3), whereas in Experiment 1 the first fixation was on the high contrast image already. This difference could be explained by the process of mental image generation, which differed between the two experiments. Indeed, in Experiment 1, participants imagined the gratings immediately after having seen the visual stimuli (short-term memory), whereas in Experiment 2, they had to create the mental images by retrieving the relevant low-level information from long-term memory. Generating the mental image required more effort and time, and only at the third fixation the high contrast information started to trigger specific eye movements. It might take about one second to deliberately generate a mental image (see Dijkstra et al. 2019 for distinctive temporal dynamics of perception vs. imagery).

Experiment 3

Task difficulty could be yet another conceivable explanation for the differences found in the first two experiments. Imagining low contrast stimuli compared to high contrast stimuli

could be more difficult, thus, participants allocated more fixations to the high contrast mental image (or vice versa). We conducted Experiment 3 to address this potential confound. We asked participants to imagine high or low contrast stimuli and to evaluate vividness of the mental image, and task difficulty. Moreover, we included a behavioral task to assure that participants did in fact create the mental images.

Method

Participants

An a priori $G \times$ Power analysis revealed that we needed a sample of 32 participants to detect a medium to large effect (paired-samples t -test, two-tailed; parameters: $d_z = 0.6$, $\alpha = 0.05$, $1 - \beta = 0.95$; Faul et al., 2007). 32 adults (24 females, aged between 19 and 35 years) participated in the experiment. All participants were naïve about the purpose of the experiment and had normal or corrected-to-normal visual acuity.

Materials and design

The task consisted of stimuli that participants had to visualize in a high contrast or a low contrast version. The stimuli were the same images used in Experiment 2 and thus consisted of 68 images representing animals like an elephant or a buffalo (34 images) or objects like a car or an umbrella (34 images). Low-level stimulus attributes were controlled via the SHINE toolbox for Matlab developed by Willenbockel et al. (2010). We matched luminance and thus minimized potential low-level confounds. The stimuli were presented in the center of the computer screen against a gray background. After each stimulus presentation, the participant visualized the image. During the visualization task, participants were shown a triangle positioned in the middle of the screen and they were instructed to create the mental image on the screen within the triangle at the exact same location as the image that was presented before. Half of the stimuli had to be visualized in a high contrast version and the other half in a low contrast version. To associate the stimuli with high and low contrast, we used upward and downward triangles. Half of the triangles during the visualization task were, thus, pointing upward and the other half downward. Participants learned the triangle-contrast association prior to the actual task. Importantly, the triangle-contrast association was counterbalanced across participants. Vividness of the mental image (1 = not vivid at all; 7 = highly vivid image) and task difficulty (1 = not difficult at all; 7 = very difficult) were assessed by means of 7-point Likert-scales. For 10 trials (out of 68), a dot appeared within the triangle before the vividness and difficulty questions and participants had to evaluate whether the dot was on their mental image or not.

The dot position was selected in a range of 30–70 pixels from the center of the screen and differed for each trial. For five trials, the dot appeared inside the prior location of the image and for five trials it was outside the prior location of the image. The dot appeared close to the border of the previously encoded image (inside or outside).

Procedure

Participants were seated in front of the computer screen. The following task instructions appeared on the screen: “You will be presented with different images (animals or objects). Your task is to visualize the images with high and low contrast. First, you will always see an image with medium contrast. Next, you will see an image with one empty triangle, in which you are going to visualize a low-contrast version or a high-contrast version of the previously seen image. If the triangle points upward, this means low contrast (*or high contrast*). If the triangle points downward, this means high contrast (*or low contrast*). When you have created the mental image, please press the space bar and answer the two subsequent questions on 7-point Likert-scales. Note that in some trials, after the mental image generation, a dot will appear before the two questions and you have to judge whether this dot is on your mental image or not. To answer this question press “f” for NO and “j” for YES”.

Learning phase Participants had to learn the association between the triangles (upward vs. downward pointing) and contrast (high vs. low contrast). Half of the participants were told that the upward pointing triangle would indicate high contrast, and a downward pointing triangle low contrast, and vice versa for the other half of participants. In the learning phase participants were presented with two stimuli and feedback was given.

Test phase In the test phase, 68 new images were presented in random order. After a fixation cross (200 ms), the image was presented for 3 s; next, a triangle pointing upward or a triangle pointing downward appeared. Half of the stimuli were followed by a triangle pointing upward and the other half by a triangle pointing downward. The order was randomized.

After each stimulus presentation, participants had to create the mental image in a high or low contrast version. As soon as they generated the mental image, they pressed the space bar (self-paced) and they had to evaluate the vividness of the image followed by task difficulty on 7-point Likert scales. Participants responded by pressing the number keys 1–7. The difficulty question always followed the vividness question (fixed order). For 10 images (5 high contrast and 5 low contrast), a dot appeared in the triangle before the vividness and difficulty questions and participants had to evaluate

whether the dot was on their mental image or not. For half of these trials the dot appeared on the prior location of the image and for the other half not.

Results

All comparisons turned out to be non-significant, thus suggesting that vividness and task difficulty were similar when it comes to imagining high vs. low contrast images. See Table 3 for means, standard deviations, and statistical analyses.

The number of correct answers in the dot task did not differ between low and high contrast conditions (see Table 3). Importantly, in both conditions, participants were above chance level. In the high contrast condition, participants correctly identified the position of 3.63 dots (mean) out of 5, which differed significantly from chance level (2.5 correct answers), $t(31)=5.64$, $p<0.001$, $d=1.00$ and in the low contrast condition, participants correctly identified the position of 3.44 dots (mean) out of 5, which also differed significantly from chance level (2.5 correct answers), $t(31)=4.80$, $p<0.001$, $d=0.85$. We, thus, conclude that participants complied with task instructions and reconstructed the mental image of the pictures previously seen.

Discussion

The results from Experiment 3 showed no differences in vividness and difficulty when imagining low vs. high contrast images. Note that null results are difficult to interpret and Likert scales might not be sensitive enough to reveal such differences. Participants complied with task instructions and imagined the previously seen pictures.

General discussion

Despite compelling empirical evidence, the role of eye movements during mental imagery is still controversially discussed. The experiments reported here were designed to investigate the influence of imagined low-level features on eye movements. Perception partly relies on low-level

features that trigger eye movements. If mental imagery shares processes with visual perception and uses depictive representations, we expect that eye movements during imagery also partly reflect properties of the imagined stimulus. In the first experiment, participants allocated more fixations to low spatial frequency–high contrast gratings not only when they saw the stimuli but also when they imagined the stimuli. This finding suggests that low-level information in mental images attract eye fixations. However, it is also possible that eye position during visual encoding led to high correspondence in eye fixations between mental imagery and visual perception. Therefore, we used images of objects (instead of gratings) that participants never saw in the low and high contrast versions (second experiment). Participants still allocated more fixations to the high contrast mental images than to the low contrast mental images. This result cannot be explained by eye position during encoding. In the third experiment, we were able to rule out task difficulty as an explanation for eye fixations toward low-level features in mental images. Taken together, we provide evidence that low-level features of mental images have a specific impact on eye movements; high contrast information in mental images triggers fixations, and this is in line with findings from visual perception. Our findings suggest that mental images embody at least partly the same pictorial features that are distinctive for the process of perception.

It has been proposed that eye movements assist the process of mental imagery (e.g., Laeng & Teodorescu, 2002). Laeng et al. (2014) offered two alternative (non-exclusive) explanations for the occurrence of eye fixations during mental imagery: (1) fixations occur because the oculomotor system does not distinguish between bottom-up driven activation of the visual cortex and top-down generated similar neuronal states, and/or (2) fixations occur as a reactivation of oculomotor memory traces. While there is partial support for the latter, the findings from our experiments offer additional support for the first explanation, suggesting that fixations occur “as a reflection of the formation of a quasi-perceptual state in the visual cortex” (Laeng et al., 2014, p. 278). There is a large amount of research investigating the neural underpinnings and this research has provided a substantial amount of evidence showing that visual perception

Table 3 Means, standard deviations (SD), paired-samples *t*-tests (*t*-values and *p*-values) and Cohen’s *d* for the vividness rating, the difficulty rating, and the number of correct answers in the dot task separated by contrast (high vs. low)

Dependent variable	Mean (SD)	<i>t</i> (31)	<i>p</i>	Cohen’s <i>d</i>
Vividness high contrast	4.99 (0.76)	1.34	0.189	0.24
Vividness low contrast	4.86 (0.96)			
Difficulty high contrast	3.16 (0.92)	0.21	0.837	0.04
Difficulty low contrast	3.17 (0.96)			
Number correct (dot task) high contrast	3.63 (1.13)	0.64	0.526	0.11
Number correct (dot task) low contrast	3.44 (1.10)			

N = 32

and mental imagery share a common neural substrate in the brain including areas that are associated with early processing of visual information (see Kosslyn & Thomson, 2003, for a meta-analysis of neuroimaging studies on mental imagery, see Pearson et al., 2015, for a review paper, see also Albers et al., 2013; Ganis et al., 2004; Lee, et al.; 2012; Slotnick et al., 2005; Stokes et al., 2009). Recent work (Bone et al., 2019), which investigated eye movements during mental imagery in relation with neural reactivation, also lends support to the pictorial account. The authors found that the specificity of neural reactivation correlated positively with fixation reinstatement, with subjective vividness of mental images (ratings) and with memory for the images (accuracy). These findings support the conceptualization of imagery processes as quasi-perceptual states. Further, it has been suggested that bottom-up processing might be involved in mental imagery (Dijkstra et al., 2019). This view proposes that during imagery early visual processing areas might be activated in a top-down manner, followed by bottom-up propagation activity (Dijkstra et al., 2019).

Earlier studies show that during mental imagery of previously seen images, participants allocate more fixations to the areas where the stimuli were encoded (Altmann, 2004; Brandt & Stark, 1997; Johansson & Johansson, 2014; Johansson et al., 2006, 2012; Laeng & Teodorescu, 2002; Laeng et al., 2014; Martarelli & Mast, 2011, 2013; Martarelli, et al., 2017; Richardson & Spivey, 2000; Scholz et al., 2016, 2018; Spivey & Geng, 2001; Wantz et al., 2016). Interestingly, these studies cannot distinguish whether the eyes return to areas where the objects were encoded, or whether low-level features in mental images are the driving force behind these eye movements. The findings from our experiments suggest that low-level features are an integral part of mental images.

Contrary to our expectations, the impact of low-level features on eye behavior during mental imagery appears to be stable over time. Effects related to the time course in both experiments were small. Considering the time course of low-level features at fixation points in Experiment 1 (see Appendix A) suggests an overall effect of the low spatial frequency–high contrast grating both in perception and imagery. Considering the time course of contrast at fixation points in Experiment 2 suggests an effect of contrast between the third and sixth fixations. However, investigating the effect further at a finer temporal resolution suggests that the effect of high contrast is not related to the temporal dynamics, but an overall effect. This is in agreement with studies in visual perception where eye fixations triggered by low-level information did not appear to change over time (Tatler et al., 2005).

There are limitations that have to be addressed. Despite the fact that participants' eyes were attracted by pictorial low-level information in mental images, there might be other

factors explaining the pattern of eye fixations. The findings are of correlational nature. On their own, they cannot confirm that low-level features are causing eye fixations during mental imagery. An interesting opportunity for future research is to include free-viewing tasks to rule out that the effects we found were influenced by the specific instructions we gave.

In conclusion, the present research provides an important contribution by showing that eye movements reflect low-level visual features of imagined stimuli. Eye movements play a central role in visual selection and construction of the perceptual world (Lauwereyns, 2012), and eye movements support the reconstruction of a perceptual world in mental imagery (Laeng & Teodorescu, 2002; Laeng et al., 2014). We propose that low-level features have a function to guide, organize, and structure mental images; this in turn might help the process of mental imagery to unfold, which is creative and flexible in kind.

Appendix A

See Fig. 4.

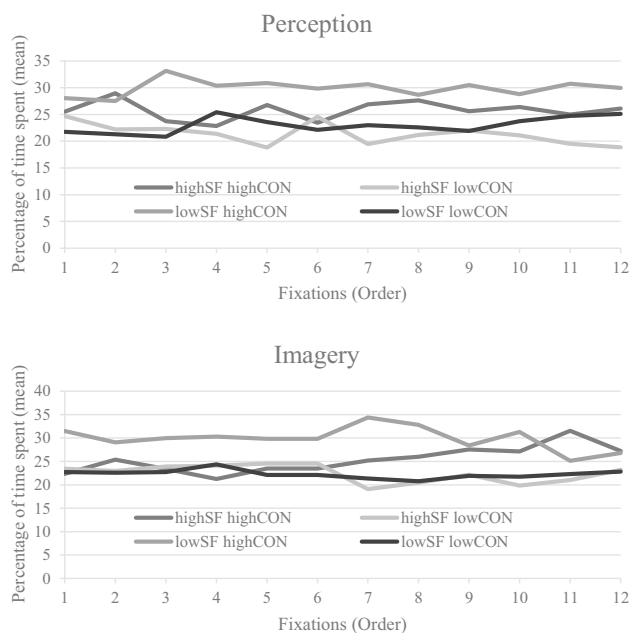


Fig. 4 Mean percentage of time spent for high contrast and low contrast visual mental images. Order of fixation is represented in the x-axis (1=first fixation until 12=12th fixation). *highSF highCON* spatial frequency high–contrast high, *highSF lowCON* spatial frequency high–contrast low, *lowSF highCON* spatial frequency high–contrast high, *lowSF lowCON* spatial frequency low–contrast low. Visual inspection of the graphs suggests an overall predominance of the *low spatial frequency–high contrast grating*, both during perception and during imagery

Appendix B

See Table 4.

Table 4 Fixation time and number of fixations for high/low spatial frequency and high/low contrast gratings separated by task (perception and imagery)

Dependent variable	Grating			
	highSF highCON	highSF lowCON	lowSF highCON	lowSF lowCON
Fixation time in ms (perception)	<i>M</i> =2251 (<i>SEM</i> =136)	<i>M</i> =2144 (<i>SEM</i> =139)	<i>M</i> =2682 (<i>SEM</i> =143)	<i>M</i> =2183 (<i>SEM</i> =153)
Number of fixations (perception)	<i>M</i> =8.22 (<i>SEM</i> =0.48)	<i>M</i> =7.95 (<i>SEM</i> =0.50)	<i>M</i> =9.92 (<i>SEM</i> =0.56)	<i>M</i> =8.07 (<i>SEM</i> =0.56)
Fixation time in ms (imagery)	<i>M</i> =1911 (<i>SEM</i> =157)	<i>M</i> =2059 (<i>SEM</i> =200)	<i>M</i> =2178 (<i>SEM</i> =183)	<i>M</i> =1806 (<i>SEM</i> =154)
Number of fixations (imagery)	<i>M</i> =4.93 (<i>SEM</i> =0.41)	<i>M</i> =4.62 (<i>SEM</i> =0.33)	<i>M</i> =5.12 (<i>SEM</i> =0.37)	<i>M</i> =4.45 (<i>SEM</i> =0.33)

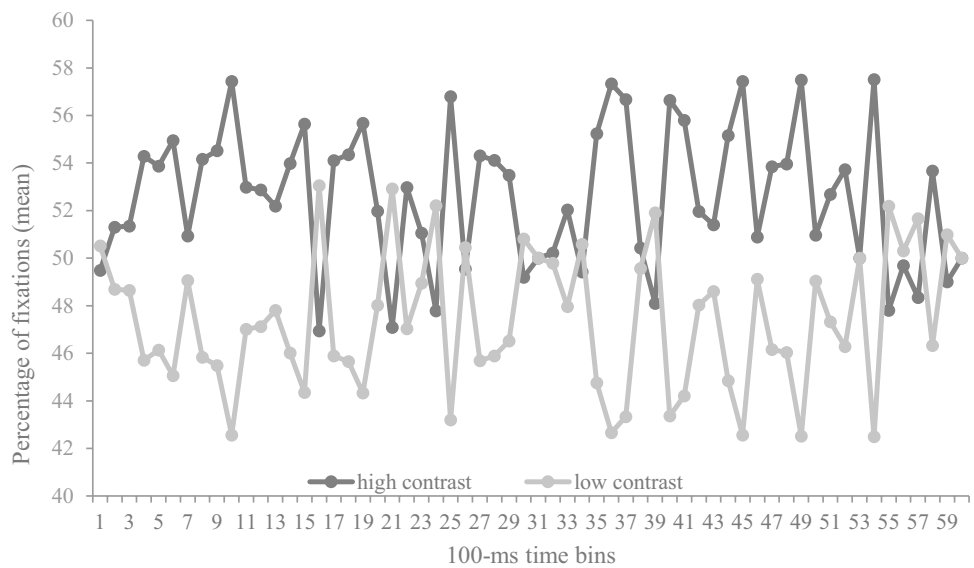
highSF highCON spatial frequency high–contrast high, *highSF lowCON* spatial frequency high–contrast low, *lowSF highCON* spatial frequency high–contrast high, *lowSF lowCON* spatial frequency low–contrast low, *M* mean, *SEM* standard error of the mean

The bold areas highlight the highest value in each row. Note that all the interactions turned out to be highly significant ($p < 0.001$) with the exception of number of fixations during imagery ($p = 0.154$).

Appendix C

See Fig. 5.

Fig. 5 Mean percentage of fixations for high contrast and low contrast visual mental images. Time (100-ms time bins) is represented in the *x*-axis (6 s stimulus presentation). Visual inspection of the graph suggests an overall predominance of the high contrast images



Acknowledgments We thank Matthias Hartmann, Felix Wichmann, the editor and two anonymous reviewers for useful comments on a previous version of the manuscript.

Funding The authors received no specific funding for this work.

Data availability statement The datasets analyzed during the current studies are available from Figsahre (<https://figshare.com/s/e64c369ef8a5dbb2ed27>).

Declarations

Competing interests The authors declare no competing interests.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study. The study (including experiments 1 and 2) was approved by the faculty ethics committee (2013-2-333838, "Augenbewegungen und Vorstellungsprozesse", Ethikkommission der philosophisch-humanwissenschaftliche Fakultät, Bern Universität).

References

- Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & de Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology*, *23*(15), 1427–1431.
- Albright, T. D. (2012). On the perception of probable things: Neural substrates of associative memory, imagery, and perception. *Neuron*, *74*(2), 227–245.
- Altmann, G. T. (2004). Language-mediated eye movements in the absence of a visual world: The “blank screen paradigm.” *Cognition*, *93*(2), B79–87.
- Baddeley, R. J., & Tatler, B. W. (2006). High frequency edges (but not contrast) predict where we fixate: A Bayesian system identification analysis. *Vision Research*, *46*(18), 2824–2833.
- Bone, M. B., St-Laurent, M., Dang, C., McQuiggan, D. A., & Ryan, J. D. B. (2019). Eye movement reinstatement and neural reactivation during mental imagery. *Cerebral Cortex*, *29*(3), 1075–1089.
- Borji, A., Sihite, D. N., & Itti, L. (2013). What stands out in a scene? A study of human explicit saliency judgment. *Vision Research*, *91*, 62–67.
- Brandt, S. A., & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience*, *9*(1), 27–38.
- Broggin, E., Savazzi, S., & Marzi, C. A. (2012). Similar effects of visual perception and imagery on simple reaction time. *The Quarterly Journal of Experimental Psychology*, *65*(1), 151–164.
- Chiquet, S., Martarelli, C.S., & Mast, F.W. (2020). Eye movements to absent objects during mental imagery and visual memory in immersive virtual reality. *Virtual Reality*.
- Dijkstra, N., Bosch, S. E., & van Gerven, M. A. J. (2019). Shared neural mechanisms of visual perception and imagery. *Trends in Cognitive Sciences*, *23*(5), 423–434.
- Einhäuser, W., & König, P. (2003). Does luminance-contrast contribute to a saliency map for overt visual attention? *European Journal of Neuroscience*, *17*(5), 1089–1097.
- Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision*, *8*(14), 1–26.
- 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*, 175–191.
- Fourtassi, M., Hajjioui, A., Urquizar, C., Rossetti, Y., Rode, G., & Pisella, L. (2013). Iterative fragmentation of cognitive maps in a visual imagery task. *PLoS ONE*, *8*(7), e68560.
- Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2004). Brain areas underlying visual mental imagery and visual perception: An fMRI study. *Cognitive Brain Research*, *20*(2), 226–241.
- Henderson, J. M., Weeks, P. A. J., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 210–228.
- Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition*, *12*(6), 1093–1123.
- Itti, L., & Borji, A. (2014). Computational models: bottom-up and top-down aspects. In A. C. Nobre & S. Kastner (Eds.), *The Oxford Handbook of Attention* (pp. 1122–1158). Oxford University Press.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10–12), 1489–1506.
- Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience*, *2*(3), 194–203.
- Johansson, R., Holsanova, J., Dewhurst, R., & Holmqvist, K. (2012). Eye movements during scene recollection have a functional role, but they are not reinstatements of those produced during encoding. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(5), 1289–1314.
- Johansson, R., Holsanova, J., & Holmqvist, K. (2006). Pictures and spoken descriptions elicit similar eye movements during mental imagery, both in light and in complete darkness. *Cognitive Science*, *30*(6), 1053–1079.
- Johansson, R., & Johansson, M. (2014). Look here, eye movements play a functional role in memory retrieval. *Psychological Science*, *25*(1), 236–242.
- Johnson, M. R., & Johnson, M. K. (2014). Decoding individual natural scene representations during perception and imagery. *Frontiers in Human Neuroscience*, *8*, 59.
- Kienzle, W., Franz, M. O., Schölkopf, B., & Wichmann, F. A. (2009). Center-surround patterns emerge as optimal predictors for human saccade targets. *Journal of Vision*, *9*(5), 1–15.
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, *4*, 219–227.
- Kosslyn, S. M., Sukel, K. E., & Bly, B. M. (1999). Squinting with the mind’s eye: Effects of stimulus resolution on imaginal and perceptual comparisons. *Memory and Cognition*, *27*(2), 276–287.
- Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The Case for Mental Imagery*. Oxford University Press.
- Laeng, B., Bloem, I. M., D’Ascenzo, S., & Tommasi, L. (2014). Scrutinizing visual images: The role of gaze in mental imagery and memory. *Cognition*, *131*(2), 263–283.
- Laeng, B., & Sulutvedt, U. (2014). The eye pupil adjusts to imaginary light. *Psychological Science*, *25*(1), 188–197.
- Laeng, B., & Teodorescu, D.-S. (2002). Eye scanpaths during visual imagery reenact those of perception of the same visual scene. *Cognitive Science*, *26*(2), 207–231.
- Lauwereyns, J. (2012). *Brain and the Gaze. On the Active Boundaries of Vision*. The MIT Press.
- Le Meur, O., Le Callet, P., & Barba, D. (2007). Predicting visual fixations on video based on low-level visual features. *Vision Research*, *47*(19), 2483–2498.

- Lee, S. H., Kravitz, D. J., & Baker, C. I. (2012). Disentangling visual imagery and perception of real-world objects. *NeuroImage*, *59*(4), 4064–4073.
- Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision*, *10*(3), 165–188.
- Martarelli, C. S., Chiquet, S., Laeng, B., & Mast, F. W. (2017). Using space to represent categories: Insights from gaze position. *Psychological Research*, *81*(4), 721–729.
- Martarelli, C. S., & Mast, F. W. (2011). Preschool children's eye-movements during pictorial recall. *British Journal of Developmental Psychology*, *29*, 425–436.
- Martarelli, C. S., & Mast, F. W. (2013). Eye movements during long-term pictorial recall. *Psychological Research*, *77*(3), 303–309.
- Naselaris, T., Olman, C. A., Stansbury, D. E., Ugurbil, K., & Gallant, J. L. (2015). A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *NeuroImage*, *105*, 215–228.
- Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, *42*(1), 107–123.
- Pearson, J., Naselaris, T., Holmes, E. A., & Kosslyn, S. M. (2015). Mental imagery: Functional Mechanisms and clinical applications. *Trends in Cognitive Sciences*, *19*(10), 590–602.
- Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition*, *76*(3), 269–295.
- Rouw, R., Kosslyn, S. M., & Hamel, R. (1997). Detecting high-level and low-level properties in visual images and visual percepts. *Cognition*, *63*(2), 209–226.
- Scholz, A., Klichowicz, A., & Krems, J. F. (2018). Covert shifts of attention can account for the functional role of “eye movements to nothing.” *Memory & Cognition*, *46*(2), 230–243.
- Scholz, A., Mehlhorn, K., & Krems, J. F. (2016). Listen up, eye movements play a role in verbal memory retrieval. *Psychological Research*, *80*(1), 149–158.
- Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception: A selective review. *Journal of Vision*, *11*(5), 1–30.
- Slotnick, S. D., Thompson, W. L., & Kosslyn, S. M. (2005). Visual mental imagery induces retinotopically organized activation of early visual areas. *Cerebral Cortex*, *15*(10), 1570–1583.
- Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research*, *65*(4), 235–241.
- Stokes, M., Thompson, R., Cusack, R., & Duncan, J. (2009). Top-down activation of shape-specific population codes in visual cortex during mental imagery. *Journal of Neuroscience*, *29*(5), 1565–1572.
- Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, *45*(5), 643–659.
- Umar, H., Mast, F. W., Cacchione, T., & Martarelli, C. S. (2021). The prioritization of visuo-spatial associations during mental imagery. *Cognitive Processing*. <https://doi.org/10.1007/s10339-020-01010-5>.
- Wantz, A. L., Martarelli, C. S., & Mast, F. W. (2016). When looking back to nothing goes back to nothing. *Cognitive Processing*, *17*(1), 105–114.
- Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior and Research Methods*, *42*(3), 671–684.
- Yarbus, A. L. (1967). *Eye Movements and Vision*. Plenum Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.