"The more you explore, the less you remember": unraveling the effects of scene clutter on learning and memory for targets

Christos Gkoumas
University of Cyprus, Silversky 3D VRT Ltd
gkoumas.christos@ucy.ac.cy

Andria Shimi
University of Cyprus
shimi.andria@ucy.ac.cy

ABSTRACT

We are constantly exposed to visually rich, oftentimes cluttered, environments. Previous studies have demonstrated the negative effects of clutter on visual search behavior and various oculomotor metrics. However, little is known about the consequences of clutter on other cognitive processes, like learning and memory. In the present study, we explored the effects of scene clutter on gaze behavior during a learning task and whether these gaze patterns influenced memory performance in a later cued recall task. Using spatial density analysis, we found that a higher degree of scene clutter resulted in more dispersed gaze behavior during the learning task. Additionally, participants recalled target locations less precisely in cluttered than in uncluttered scenes during the recall task. These findings have important implications for theories linking exploratory viewing with memory performance as well as for making recommendations on how interior spaces could be better organized to facilitate daily living.

CCS CONCEPTS

• **Applied computing** → Law, social and behavioral sciences; Psychology; • **Human-centered computing** → Human computer interaction (HCI); Empirical studies in HCI.

KEYWORDS

clutter, learning, memory, eye tracking, spatial density (gaze entropy)

ACM Reference Format:

Christos Gkoumas and Andria Shimi. 2022. "The more you explore, the less you remember": unraveling the effects of scene clutter on learning and memory for targets. In *2022 Symposium on Eye Tracking Research and Applications (ETRA '22), June 08–11, 2022, Seattle, WA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3517031.3529623

1 INTRODUCTION

The visual characteristics of our surrounding environment play a major role in how we perceive and process information [Rodrigues and Pandeirada, 2020]. One such characteristic is visual clutter (VC), which refers to a state of an excessive number of poorly organized objects in a scene [Soojin et al., 2015]. VC can be found in a wide range of interior spaces (e.g., classrooms, offices, living rooms, etc.) with a detrimental impact on critical cognitive processes, such as visual search, and several eye movement metrics. Therefore, a better understanding of the functional consequences of VC on cognitive functioning and eye movements is of fundamental importance to guiding everyday actions and natural behavior.

The effects of VC have been studied more thoroughly in the context of visual search. Previous studies have demonstrated that VC increases search times and error rates in visual search tasks using images of real-world scenes [Asher et al., 2013; Henderson et al., 2009], images of different types of cities [Neider and Zelinsky, 2011], images of bags with personal belongings [Bravo and Farid, 2008], simulated graphics programs [Moacdieh and Sarter, 2017] and aeronautical charts [Beck et al., 2010; Beck et al., 2012]. Apart from these decrements in behavioral performance, VC has also been shown to influence several eye movement metrics. Specifically, participants tend to make more fixations [Beck et al., 2010; Beck et al., 2012], have longer decision times (total search time – time to the first fixation to the target) [Neider and Zelinsky, 2011] and fixation durations [Delmas et al., 2022; Henderson et al., 2009] as well as move their eyes less directly to the target, as indicated by scan path ratios [Neider and Zelinsky, 2011] when searching under highly cluttered compared to uncluttered conditions. In sum, these findings suggest that VC affects visual search performance broadly, both in terms of behavioral and eye movement metrics.

Although previous studies have demonstrated the detrimental impact of VC on visual search behavior, there are a few issues that have not yet been addressed. One issue pertains to the fact that most VC studies have employed one-off visual search tasks, i.e., tasks that participants had to search for the target only once. What happens in cases of repeated exposure to VC? In one study, [Raines et al., 2014] tested people with hoarding disorder, a condition associated with the acquisition and storing of an excessive number of items in a chaotic manner, with a sustained attention task and a verbal learning task in a cluttered and non-cluttered room. They found that task performance did not differ between the two clutter conditions, thus arguing that repeated exposures to cluttered environments due to hoarding may have resulted in habituation, which in turn attenuated the negative effects of clutter on cognitive processes. In another study, [Võ and Wolfe, 2013] asked participants to perform a repeated visual search task where they had to find objects that were placed in semantically-inconsistent locations within scenes (a condition that may mimic a state of VC). Participants became faster at finding the target object as blocks progressed, which the authors attributed to the greater involvement of episodic memory to guide the search process when no semantic information was available. Nevertheless, this study aimed to examine the conditions under which different types of memory (semantic vs. episodic) guided repeated search, still, leaving unanswered how well participants

ETRA '22, June 08–11, 2022, Seattle, WA, USA

ACM ISBN 978-1-4503-9252-5/22/06.
https://doi.org/10.1145/3517031.3529623

learn target locations in scenes with different levels of VC, to which they have been repeatedly exposed.

A second related issue is that memory performance for previously-searched targets has not been directly examined in previous VC studies. Therefore, measuring the effects of VC on processes other than visual search, such as the formation of long-term memory representations, can extend significantly our knowledge on this topic.

A third issue is related to the underutilization of specific eye-tracking metrics in the VC literature. One such metric is spatial density [also referred to as stationary gaze entropy, Shiferaw et al., 2019], i.e., the spatial dispersion of fixations on a display or interface [Goldberg and Kotval, 1999; Moacdieh and Sarter, 2015]. A pattern of fixations spread all over the display indicates greater exploratory viewing resulting from extensive processing of information. Contrary, fixations covering a small area of the display are thought to reflect a structured and more directed path towards a potential target. In a recent study, [Moacdieh and Sarter, 2017] investigated the effects of two core components of display clutter, i.e., data density and organization, on spatial density in a simulated graphics program. They found that increases in data density and low levels of organization led to more dispersed gaze behavior, thus rendering the use of spatial density analysis a sensitive and reliable measure of clutter [see also Kanaan and Moacdieh, 2021]. In addition, spatial density has been used as an index of task load, with increased load associated with more diffused allocation of attention [Di Stasi et al., 2016]. To our knowledge, spatial density analysis has not been used in studies with naturalistic images of scenes and manipulations of VC.

To address these issues, we employed two tasks, namely a search-based learning (SBL) and a cued recall (CR) task and recorded participants' eye movements while carrying out the tasks. In the SBL task, participants were asked to detect and explicitly memorize the location of targets presented within cluttered and uncluttered scenes over four learning blocks. After a delay, their long-term memory for these locations was probed in a CR task. To obtain a more thorough understanding of how VC influences SBL, we used three eye-tracking metrics. Primarily, we were interested in fixation spatial density (spread metric), which reflected the degree of visual exploration of scenes by capturing how dispersed fixation locations were. Secondarily, we measured average saccade length (directness metric), which captured characteristics of the sequence of fixation locations. Finally, we measured fixation durations (duration metric), as an index of in-depth processing of information [Gameiro et al., 2017; for a review of these metrics see Moacdieh and Sarter, 2015]. Additionally, we examined the relationship between these metrics during the SBL task and the recall performance for target locations in the CR task. This relation was important to address, especially in light of recent findings suggesting that greater visual exploration during scene encoding was linked to better performance in a subsequent memory task [Fehlmann et al., 2020; Ramey et al., 2020].

In the SBL task, we hypothesized that fixations will be (1) more dispersed, indicating exhaustive visual exploration, (2) closer to each other, and (3) last longer in cluttered than in uncluttered scenes. Fourthly, we predicted that, in the CR task, memory for target locations will be equally precise in both clutter scenes, based on recent findings indicating that once participants have learned the location of targets, VC does not influence memory performance [Gkoumas and Shimi, 2021]. Lastly, we expected gaze dispersion patterns during learning to be strongly associated with memory performance in the CR task.

2 METHODS

2.1 Participants

A total of 29 participants (21 females, M_{age}= 20.9 ± 4.44 years, five left-handed) aged 18 to 38 years old were recruited from the Department of Psychology of the University of Cyprus. Participants had normal or corrected-to-normal vision and reported no history of neurological or psychiatric impairments. All participants gave informed consent before participating in the study and received course credit for their participation. The study was approved by the Cyprus National Bioethics Committee.

2.2 Apparatus and Procedure

2.2.1 ***Equipment.*** Eye movements were collected using a GP3 HD eye tracker (Gazepoint Research Inc., Canada) with a sampling frequency of 150 Hz and with a reported accuracy of 0.5-1° of visual angle. Participants sat 65 cm away from a Philips 223V monitor (21.5 inches, 60 Hz) with a screen resolution of 1920 x 1080 pixels. To avoid unnecessary head movements during the experimental session, participants were placed in a chin and forehead rest.

2.2.2 ***Stimuli.*** Study materials included 72 images of indoor scenes (1920 x 1080 pixels) whereas five additional images were used in practice trials at the beginning of each task. Image classification into cluttered (hereafter high clutter) or uncluttered (heareafter low clutter) scenes was based on a two-stage process, including subjective ratings of clutter by human observers and the well-established computational measure of edge density (see *Supplementary Material* for a full description of the clutter rating procedure). All images illustrated scenes and did not include any people or animals to avoid interference from distracting stimuli. The target object in the SBL task was a small golden star (20 x 20 pixels). This semantically meaningless target was used to ensure that its location could not be predicted by contextual scene-related information.

2.2.3 ***Experimental Tasks.*** Participants completed two tasks, namely a search-based learning and a cued recall task.

Search-based learning task. In this task, participants had to detect and explicitly memorize the location of a target star superimposed on cluttered and uncluttered scenes. Participants were presented with 72 images of naturalistic scenes. These were presented sequentially and in random order within a block and were repeated over four blocks. Each trial began with a fixation cross at the center of the screen (black cross in grey background), which participants fixated for 500 milliseconds (ms). Subsequently, a new scene appeared on the screen and participants had 35 seconds maximum to detect and learn the unique location of a target star for an upcoming memory test. Once they detected the target star, they indicated its location by clicking on it using the mouse while fixating their gaze on the star. Their response was marked as accurate if their click and gaze (at the time of response) were within 50 and 100 pixels around the target respectively. In that case, a feedback message (green checkmark in grey background) appeared on the screen for 700 ms. If they clicked outside this boundary or if the time elapsed, the appropriate message was displayed depending on the condition (a red X or a clock, respectively), and their response was considered incorrect. In half the images (36) the star was located on the right side of the scene (18 top right, 18 bottom right), and in the remaining half, it was located on the left side (18 top left, 18 bottom left). The location of the star was counterbalanced across scenes and participants. Participants were given short self-paced breaks every 24 trials and at the end of each block (72 trials), at which they were informed about their accuracy (i.e., the number of stars they had collected). A built-in 9-point calibration procedure was run before the main task and was repeated at the end of each block if needed. A calibration score was only accepted if the average calibration error was below 50 pixels or 1° degree of visual angle.

Cued recall task. In this task, participants were presented with the same scenes they studied during the SBL task, but this time without the target star. Each trial began with a fixation cross at the center of the screen, which participants fixated for 500ms to proceed. Then, a previously-studied scene appeared, and participants had 35 seconds to recall the target location and indicate it using the mouse. After each response, at the end of each trial, participants had 20 seconds to rate their confidence, on a 5-point scale (1=Not at all, 2=Slightly, 3=Moderately, 4=Fairly, 5= Completely), regarding the star location they indicated. Performance in the CR task was measured using memory precision, which is the Euclidean distance (in pixels) between the recalled location indicated by each participant and the actual location of the star in the SBL task. Calibration procedures and scores were similar to those in the SBL task.

2.2.4 ***Procedure.*** Participants were tested individually in a dimly lit laboratory room at the University of Cyprus. At the beginning, the examiner introduced participants to the study procedures and explained the use and purpose of the chin and forehead rest. Before each task, the examiner explained the trial sequence using printed screenshots of an example trial. Then, the height of the chin and forehead rest was adjusted to fit the participant's needs and the experimenter ensured that the participant had a comfortable seat. A custom 9-point calibration routine preceded each task until an acceptable calibration error score was obtained. The calibration routine was also executed during the SBL task if needed. Given the goals of the study, participants first carried out the SBL task and then the CR task, with a 10-minute break in-between.

2.3 Data Pre-processing

Our preprocessing pipeline was executed in four steps [Duchowski, 2017]. The first step involved discarding missing points from raw data (e.g., eye blinks, gaze coordinates out of screen range). Second, we computed the Euclidean distance between consecutive gaze samples (for x and y gaze coordinates) using $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$, where (x_1, y_1) and (x_2, y_2) represented the pixel coordinates of two consecutive gaze samples. Third, the time difference between consecutive samples was computed by subtracting the timestamp of the previous sample from that of the current sample. Gaze velocities, θ, were computed using the distance traversed by the eyes divided by the sample-to-sample change in time. Finally, we converted velocities from pixels/s to degrees of visual angle (°)/s.

2.3.1 ***Filtering and denoising.*** Eye-tracking data corresponding to unwanted recorded segments (breaks, practice trials, eye blinks, out of range values) were discarded before proceeding to event detection. Samples, where gaze velocity exceeded 1000°/s, were also removed as they probably represented unphysiologically fast eye movements [Nyström and Holmqvist, 2010]. Finally, a second-order polynomial Savitzky-Golay filter with a width of 11 samples (~73 ms) was applied to gaze velocity data to smooth the signal and account for noise before event detection.

2.3.2 ***Event Detection algorithm.*** We used the Velocity-Threshold Identification (I-VT) algorithm to categorize gaze samples as fixation or saccades [Salvucci and Goldberg, 2000]. According to the I-VT algorithm, gaze samples below a certain velocity threshold are marked as fixation points, otherwise, they are considered to be part of a saccade. Consecutive fixation points are grouped into a fixation group, whose centroid is the mean of all points in the same fixation group. In our study, gaze samples with velocities below 120°/s were classified as fixations. A minimum duration of 50 ms was selected for fixations [Nuthmann, 2017].

2.4 Spatial Density Analysis

In the present study, we used spatial density analysis to examine how different levels of scene clutter influence fixation spatial distribution patterns [Goldberg and Kotval, 1999; Moacdieh and Sarter, 2015]. We divided each image into a 15x10 grid (150 grid cells in total), with each grid cell covering 128x108 pixels, as shown in Figure 1. The spatial density score was then calculated as the percentage (%) of the total number of grid cells that contained at least one fixation point divided by the total number of grid cells.

3 RESULTS

Data analyses were performed with custom-made scripts in Python. Unless otherwise stated, the data were analyzed using a two-way repeated-measures Analysis of Variance (RM ANOVA) with learning blocks (Block 1, 2, 3, 4) and scene clutter level (uncluttered vs. cluttered) as independent factors. Assumption of normality was assessed using the Shapiro-Wilk test, visual inspection of histograms and Q-Q plots, and by calculating the ratios of skewness and kurtosis to their standard errors, respectively. Greenhouse-Geisser correction is reported in case of sphericity violations. Bonferroni correction for multiple comparisons was used when necessary. Eye tracking data from one participant were not stored properly due to technical issues, therefore we report findings from 28 participants unless otherwise specified. In the SBL task, scenes in which participants detected the target in less than two of the four learning blocks (<50%) were discarded, as it cast doubts on how well they learned the location of the target (scenes excluded: 1.57%). From the remaining scenes, only those in which participants detected the target successfully were included in the analyses presented here. For example, if a target was detected in three out of four learning blocks for scene number 25, the scene was included in the analysis but only the three successful blocks were taken into account and the other one was excluded (trials excluded: 1.84%).

Figure 1: Examples of spatial density analysis for an uncluttered (left) and a cluttered (right) scene used in our study. White grids contained at least one fixation. On the left uncluttered image, 11 grids were fixated resulting in a spatial density score of 7.3% (11*100/150). On the right cluttered image, the spatial density score was 16% (24*100/150). Grid cells containing the letter T correspond to the location of the target in the scene for a participant. Left image via pxfuel.com, right image "2012-145 My Messy Room" by Denise Krebs is licensed under CC BY 2.0/Gridded from original.

3.1 Search-based Learning task

Initially, we conducted an RM ANOVA with fixation spatial density as the dependent variable. Results showed that overall, participants had more dispersed gaze in cluttered (*M*= 7.65% , *SD*= 2.30%) than in uncluttered scenes (*M*= 4.61% , *SD*= 1.44%), *F*(1,27)=237.52, *p*<.001, $\eta^2{}_p$=.897. Furthermore, gaze patterns became less dispersed as the learning blocks progressed *F*(1.90,51.46)=70.65, *p*<.001, $\eta^2{}_p$=.723 (Block 1: 7.64% , Block 2: 6.65% , Block 3: 5.59% , Block 4: 4.64% , all post hoc comparisons were p<.001). These main effects were also accompanied by a significant clutter level x learning block interaction *F*(3,81)=8.59, *p*<.001, $\eta^2{}_p$=.241. Post-hoc comparisons indicated that the interaction was driven by diminishing differences in spatial density between the two clutter conditions over blocks. Figure 2 shows the results.

We then extracted regression slopes for all participants for the two clutter levels. Here, slopes represented how much spatial density improved over the blocks in each clutter level or, in other words, how fixation spatial distribution became less dispersed. The analysis of slopes revealed that participants showed greater improvement in spatial density over the blocks in cluttered (*M*= -1.23, *SD*=0.69) than in uncluttered scenes (*M*= -0.78, *SD*=0.48), *t* (27) = 4.41, *p*<.001 (the more negative the slope the greater the improvement over blocks).

Following spatial density analyses, we examined two other variables of interest, i.e., average saccade length (in pixels) and average fixation durations. Average saccade length indicates the distance traveled by the eyes considering multiple fixation points, whereas average fixation duration indicates the time during which the eyes fixated on the same or relatively the same location while participants viewed the scene. Results from the RM ANOVA, with average saccade length as the dependent variable, showed that participants made significantly longer saccadic movements in cluttered (*M*= 321 pixels, *SD*= 49 pixels) than in uncluttered scenes (*M*= 311 pixels, *SD*= 54 pixels), *F*(1,27)=13.87, *p*<.001, $\eta^2{}_p$=.339. In addition, the average saccade length decreased significantly over blocks (*F*(1.97,53.21)=5.12, *p*=.009, $\eta^2{}_p$=.159), from 325 pixels (*SD* =49 pixels) in Block 1 to 321 pixels (*SD* = 46 pixels) in Block 2, 313 pixels (*SD* = 52 pixels) in Block 3 and 306 pixels (*SD* = 56 pixels)

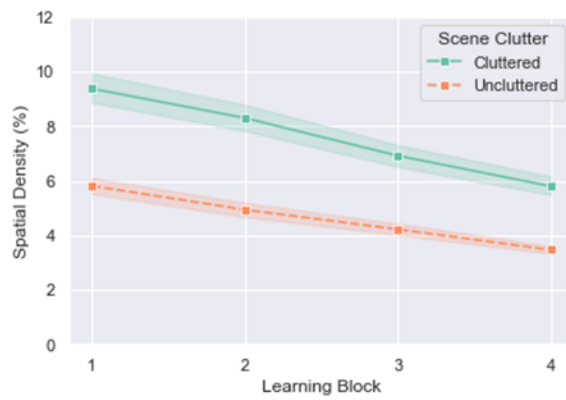

Figure 2: Spatial density (%) in each learning block for cluttered (solid line) and uncluttered (dashed line) scenes. Ribbons represent 95% confidence intervals.

in Block 4 (all post-hoc comparisons were significant, p<0.05). The interaction between scene clutter and learning blocks for average saccade length did not reach significance (p>0.05). Finally, results from the RM ANOVA with average fixation durations as the dependent variable (one participant had consistently outlying values and was excluded from this analysis) revealed only a main effect of scene clutter, *F* (1,26) = 59.81, *p*<0.001, $\eta^2{}_p$=.697, with longer fixations in uncluttered (*M*= 335 milliseconds, *SD*=79.2 milliseconds) than in cluttered scenes (*M*= 301 milliseconds, *SD*=61.7 milliseconds), which was a counterintuitive finding.

3.2 Cued Recall task

A paired-sample t-test was conducted to examine the effects of scene clutter level on memory precision. Results showed that participants recalled target locations more precisely in uncluttered (*M*= 194 pixels, *SD*= 147.1 pixels) than in cluttered scenes (*M*= 230.8 pixels, *SD*= 160.6 pixels), *t*(27)=3.22, *p*=.003, *d* =.61. Since memory precision represented the distance of the participant's response from the target location in pixels, smaller values of precision indicate that participants had better memory representations of the target location.

To further investigate the relationship between eye movement metrics (i.e., spatial density, average saccade length, average fixation duration) during the SBL task and recall of target locations during the CR task, we fitted two multiple regression models. For uncluttered scenes, spatial density in the last learning block (4^{th}) and the slope of spatial density across blocks explained 54% of the variance of recall performance. The multiple regression model was statistically significant, $F(2, 25) = 16.44$, $p < .001$, adj. $R^2 = .54$, with both variables having statistically significant contribution to the prediction ($p<0.001$). For cluttered scenes, spatial density in the last learning block (4^{th}) and the slope of spatial density across blocks predicted recall performance significantly, $F(2, 25) = 15.25$ $p < .001$, adj. $R^2 = .51$. The two variables added significantly to the prediction ($p<0.001$).

4 DISCUSSION

The goals of this study were to examine the impact of scene clutter on gaze behavior during an SBL task and on recall performance for target locations in a subsequent CR task. In addition, we were interested in identifying eye gaze markers during the SBL task that predicted memory performance in the CR task, thus providing new insights on howgaze behavior at encoding under differential VC predicts performance at retrieval. To this end, fixation spatial density, average saccade length, and average fixation durations were measured while participants searched and memorized the location of a target within cluttered and uncluttered scenes over multiple blocks. We then examined the extent to which these three metrics explained the observed performance in a CR task for target locations in the previously-studied scenes.

Our *first hypothesis* suggested that fixations will be more dispersed in cluttered than in uncluttered scenes during the SBL task. Using spatial density analysis, we found that gaze dispersion was indeed greater in cluttered than in uncluttered scenes. This finding is in line with previous VC studies that have reported that spatial density is a sensitive measure to capture the effects of clutter on search performance [Kanaan and Moacdieh, 2021; Moacdieh and Sarter, 2017]. Here, we extend past findings by showing the effects of VC on spatial density using complex images of naturalistic scenes, and a different experimental paradigm, in which participants had to memorize the location of targets. The dense amount of information in cluttered scenes likely promotes a more explorative viewing behavior compared to the information sparsity observed in uncluttered scenes [Shiferaw et al., 2019]. Moreover, we observed that the difference in spatial density between cluttered and uncluttered scenes remained over blocks, despite getting smaller. This finding suggests that changes in spatial density are sensitive both to the amount of scene clutter and to the number of times exposed to the same scenes. Although previous research in adults with hoarding disorder demonstrated that information processing is largely independent of the level of clutter in the environment [Raines et al., 2014], our findings suggest that this is not the case for healthy young adults.

Our *second hypothesis* posited that fixations will be closer to each other in cluttered than in uncluttered scenes, as indicated by shorter average saccade length in cluttered than uncluttered scenes. However, we found that participants made, on average, longer saccadic movements in cluttered than uncluttered scenes. Previous research on the effects of VC on mean saccade length has provided mixed findings. For instance, [Henderson et al., 2009] found no effects of clutter on mean saccade length in a visual search task using scene photographs. In contrast, [Moacdieh and Sarter, 2017] reported that participants made longer saccades while searching low density, well-organized displays in a simulated graphics program compared to high density, poorly-organized displays. One potential explanation for our results may relate to the nature of the task, that is, participants had to detect and memorize the target locations for an upcoming memory test instead of simply searching for them as in past VC studies. It may be the case that participants strategically made longer saccadic movements in cluttered scenes as this would help them scan as much of the scene as possible and therefore increase their chances of detecting the target and memorizing its location for the upcoming memory test.

Contrary to our *third hypothesis* about average fixation durations, we found that fixations lasted longer in uncluttered than in cluttered scenes. It is important to note that this difference between the two clutter levels did not change over blocks, providing evidence that it may reflect baseline differences in how participants sampled information under the two clutter levels. However, this finding stands in stark contrast to a well-established and widely replicated finding in VC literature, which suggests that higher levels of clutter prolong fixation durations [Delmas et al., 2022; Henderson et al., 2009; Moacdieh and Sarter, 2017]. Thus, what may explain the conflicting finding between our study and past studies? In our view, it seems possible that task demands influenced participants’ fixation duration [Nuthmann, 2017]. As with average saccade length, participants may have found it costly to fixate for long in a few locations only, given the amount of to-be-scanned information in a cluttered scene, and opted instead for shorter fixations, as this would allow them to scan a much wider proportion of a cluttered scene, and therefore maximize their chances of detecting the target.

Taken together, the analysis of the three eye movement metrics has yielded new insights regarding the strategies that participants used during the SBL task. Specifically, participants explored cluttered scenes more extensively (greater gaze dispersion), their fixations were further from each other (longer average saccade length) and shorter in duration (shorter fixation durations) compared to uncluttered scenes. These results suggest that participants favored widespread scanning at the expense of scanning carefully under highly cluttered conditions. This pattern reflects the classic exploration-exploitation dilemma, according to which the time allocated for exploration limits the time available for scrutinizing certain locations [Berger-Tal et al., 2014; Cohen et al., 2007; Gameiro et al., 2017]. Further support for this trade-off comes froma closer inspection of the average saccade lengths and fixation durations that participants made, in line with the ambient-focal visual scene processing hypothesis. Based on this hypothesis, short fixation durations followed by large saccades are indicative of ambient scanning (exploration), whereas long fixations followed by short saccades are indicative of focal scanning (inspection) [Krejtz et al., 2016; Unema et al., 2005]. Thus, our results indicate that in cluttered scenes, participants adopted an ambient scanning strategy (extensive exploration), while in uncluttered scenes they used focal scanning, which has been linked to more focused viewing patterns.

Next, we examined the quality of memory representations that participants formed after learning the location of targets within cluttered/uncluttered scenes. Interestingly, participants formed more precise memories of target locations in uncluttered than in cluttered scenes during the CR task. The findings from the SBL task in conjunction with those in the CR task suggest that ambient scanning (exploration) of cluttered scenes during SBL resulted in less precise memories for target locations, whereas focal scanning (inspection) of uncluttered scenes resulted in more precise memory representations. This pattern of results adds significantly to the literature [Fehlmann et al., 2020; Ramey et al., 2020] as it shows that exploratory scanning of a scene during encoding relates to poorer memory performance at recall, and highlights scene clutter as a potential factor that can alter exploratory gaze behavior during learning with downstream effects in the quality of long-term memory representations. The underlying process supporting our finding may be that greater exploration in cluttered scenes in the SBL task reflects participants' effort to resolve the clutter and detect the target and is not necessarily a sign of more distributed attentional allocation in favor of sampling more visual information for the upcoming memory task. Further research is needed to examine this hypothesis.

Finally, we looked at the relation between the three eye movement metrics we measured in the SBL task and memory precision for target locations in the CR task. The best prediction models revealed that the slope of spatial density across blocks, that is, how gaze dispersion improved over the blocks, and the spatial density in the last learning block (4^{th}) were the only significant predictors of memory precision in both clutter scenes. More specifically, the less scattered the fixation locations became over blocks and the less dispersed they were in the last learning block, the more precise the memory representations were. Importantly, these two predictors accounted for more than half (51% in cluttered, 54% in uncluttered) of memory precision variance. Our findings emphasize the predictive power of spatial density and identify it as a strong candidate for tracing the effects of VC (or complexity, in general) on long-term memory in naturalistic complex environments, such as images of scenes, Virtual/Augmented Reality applications or real-world scenarios, using eye-tracking [for a similar application of this metric in a simulated procedure, see Diaz-Piedra et al., 2017; for recent applications using gaze entropy in Virtual Reality, see Harris et al., 2021; Chung et al., 2022].

5 CONCLUSION

VC is an inherent attribute of the environments that we live in. We manipulated the amount of VC in images of naturalistic scenes to track down its effects on learning and later recalling of target locations, as well as on certain eye movement metrics. The present study is, to the best of our knowledge, the first to show that VC in images of complex, naturalistic scenes influences how people search for targets embedded in these scenes and negatively impacts the quality of the representations formed in long-term memory. Specifically, our results suggest that VC influences the long-term learning of target locations by affecting primarily the spatial dispersion of gaze (spread metric) and secondarily the way participants look at certain locations to process information (directness and duration metrics). These eye gaze markers are therefore particularly helpful to be used in studies aiming to understand further how VC may influence cognition. Indeed, here we identified aspects of fixation spatial density during learning as significant predictors of memory performance, further emphasizing the value of this underutilized eye tracking metric. Future applications can capitalize on our finding by examining the usefulness of this metric in the context, for example, of simulation-based skill assessment and training programs, where both the amount of information available and users' gaze behavior are of primary interest, to maximize learning outcomes. Finally, our findings showed that participants formed more precise memories for previously studied targets in uncluttered than in cluttered scenes. From a practical standpoint, our findings indicate that decluttering our surrounding environment might alleviate the detrimental impact of VC on cognitive processes like learning and memory and promote a more cognitive-friendly way of living.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Innovative Training Network, Grant Agreement No 813546.

REFERENCES

Matthew F Asher, David J Tolhurst, Tom Troscianko, and Iain D Gilchrist. 2013. Regional effects of clutter on human target detection performance. *Journal of vision* 13, 5 (2013), 25-25.

Melissa R Beck, Maura C Lohrenz, and Gregory Trafton. 2010. Measuring search efficiency in complex visual search tasks: Global and local clutter. *Journal of experimental psychology: applied* 16, 3 (2010), 238.

Melissa R Beck, Michael Trenchard, Amanda Van Lamsweerde, Rebecca R Goldstein, and Maura Lohrenz. 2012. Searching in clutter: Visual attention strategies of expert pilots. In *Proceedings of the Human Factors and Ergonomics Society annual meeting*, 1411-1415.

Oded Berger-Tal, Jonathan Nathan, Ehud Meron, and David Saltz. 2014. The exploration-exploitation dilemma: a multidisciplinary framework. *PloS one* 9, 4 (2014), e95693.

Mary J Bravo and Hany Farid. 2008. A scale invariant measure of clutter. *Journal of Vision* 8, 1 (2008), 23-23.

Jiyong Chung, Hyeokmin Lee, Hosang Moon, and Eunghyuk Lee. 2022. The Static and Dynamic Analyses of Drivers' Gaze Movement Using VR Driving Simulator. *Applied Sciences* 12, 5 (2022), 2362.

Jonathan D Cohen, Samuel M McClure, and Angela J Yu. 2007. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences* 362, 1481 (2007), 933-942.

Maxime Delmas, Loïc Caroux, and Céline Lemercier. 2022. Searching in clutter: Visual behavior and performance of expert action video game players. *Applied Ergonomics* 99 (2022), 103628.

Leandro L Di Stasi, Carolina Diaz-Piedra, Héctor Rieiro, Jose M Sanchez-Carrion, Mercedes Martin Berrido, Gonzalo Olivares, and Andrés Catena. 2016. Gaze entropy reflects surgical task load. *Surgical endoscopy* 30, 11 (2016), 5034-5043.

Carolina Diaz-Piedra, Jose M Sanchez-Carrion, Héctor Rieiro, and Leandro L Di Stasi. 2017. Gaze-based technology as a tool for surgical skills assessment and training in urology. *Urology* 107 (2017), 26-30.

Andrew T Duchowski. 2017. *Eye tracking methodology: Theory and practice*. Springer.

Bernhard Fehlmann, David Coynel, Nathalie Schicktanz, Annette Milnik, Leo Gschwind, Pascal Hofmann, Andreas Papassotiropoulos, and Dominique JF de Quervain. 2020. Visual exploration at higher fixation frequency increases subsequent memory recall. *Cerebral cortex communications* 1, 1 (2020), tgaa032.

Ricardo R Gameiro, Kai Kaspar, Sabine U König, Sontje Nordholt, and Peter König. 2017. Exploration and exploitation in natural viewing behavior. *Scientific Reports* 7, 1 (2017), 1-23.

Christos Gkoumas and Andria Shimi. 2021. Levels of visual clutter differentially impact search-based learning in naturalistic scenes. *Journal of Vision* 21, 9 (2021), 2081-2081.

Joseph H Goldberg, and Xerxes P Kotval. 1999. Computer interface evaluation using eye movements: methods and constructs. *International journal of industrial ergonomics* 24, 6 (1999), 631-645.

David J Harris, Kyle J Hardcastle, Mark R Wilson, and Samuel J Vine. 2021. Assessing the learning and transfer of gaze behaviours in immersive virtual reality. *Virtual Reality* 25, 4 (2021), 961-973.

John M Henderson, Myriam Chanceaux, and Tim J Smith. 2009. The influence of clutter on real-world scene search: Evidence from search efficiency and eye movements. *Journal of vision* 9, 1 (2009), 32-32.

Malk Kanaan and Nadine Marie Moacdieh. 2021. How do we react to cluttered displays? Evidence from the first seconds of visual search in websites. *Ergonomics* 64, 11 (2021), 1452-1464.

Krzysztof Krejtz, Andrew Duchowski, Izabela Krejtz, Agnieszka Szarkowska, and Agata Kopacz. 2016. Discerning ambient/focal attention with coefficient K. *ACM Transactions on Applied Perception (TAP)* 13, 3 (2016), 1-20.

Nadine Marie Moacdieh, and Nadine Sarter. 2015. Display clutter: A review of definitions and measurement techniques. *Human factors* 57, 1 (2015), 61-100.

Nadine Marie Moacdieh and Nadine Sarter. 2017. The effects of data density, display organization, and stress on search performance: An eye tracking study of clutter. *IEEE Transactions on Human-Machine Systems* 47, 6 (2017), 886-895.

Mark B Neider, and Gregory J Zelinsky. 2011. Cutting through the clutter: Searching for targets in evolving complex scenes. *Journal of Vision* 11, 14 (2011) 7-7.

Marcus Nyström and Kenneth Holmqvist. 2010. An adaptive algorithm for fixation, saccade, and glissade detection in eye tracking data. *Behavior research methods* 42, 1 (2010), 188-204.

Antje Nuthmann. 2017. Fixation durations in scene viewing: Modeling the effects of local image features, oculomotor parameters, and task. *Psychonomic bulletin & review* 24, 2 (2017), 370-392.

Amanda M Raines, Kiara R Timpano, and Norman B Schmidt. 2014. Effects of clutter on information processing deficits in individuals with hoarding disorder." *Journal of affective disorders* 166 (2014), 30-35.

Michelle M Ramey, John M Henderson, and Andrew P Yonelinas. 2020. The spatial distribution of attention predicts familiarity strength during encoding and retrieval." *Journal of Experimental Psychology: General* 149, 11 (2020), 2046.

Pedro FS Rodrigues and Josefa NS Pandeirada. 2020. The influence of the visual surrounding environment in older adults and young adults' cognitive performance: an alternative paradigm. *Journal of Cognitive Psychology* 32, 3 (2020), 332-343.

Dario D Salvucci and Joseph H Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*. 71–78.

Brook Shiferaw, Luke Downey, and David Crewther. 2019. A review of gaze entropy as a measure of visual scanning efficiency. *Neuroscience & Biobehavioral Reviews* 96 (2019), 353-366.

Park Soojin, Talia Konkle, and Aude Oliva. 2015. Parametric coding of the size and clutter of natural scenes in the human brain. *Cerebral cortex* 25, 7 (2015), 1792-1805.

Pieter JA Unema, Sebastian Pannasch, Markus Joos, and Boris M Velichkovsky. 2005. Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual cognition* 12, 3 (2005), 473-494.

Melissa L-H Võ, and Jeremy M Wolfe. 2013. The interplay of episodic and semantic memory in guiding repeated search in scenes. *Cognition* 126, 2 (2013), 198-212.