Access and Use of Contextual Expectations in Visual Search during Aging

Miguel Borges (m.borges@student.maastrichtuniversity.nl)[1]
Moreno I. Coco (micoco@psicologia.ulisboa.pt)
University of Lisbon, Department of Psychology,
Alameda da Universidade, 1600-214 Lisboa, Portugal

Abstract

Expectations are routinely utilized by the cognitive system to facilitate processing and optimize responses. However, as we age, the access and use of expectations might undergo significant changes. In an eye-tracking study, we investigate how search performance in young and old age is influenced by contextual priming of expectations, and the contextual consistency of the target object. We demonstrate that older participants congruently primed perform significantly worse, especially when the target object is inconsistent with the scene context (e.g., exposed to a 'restaurant' scene before performing the search in another 'restaurant' scene, but looking for an 'iron' rather than 'bread'). Even though congruency and consistency differentially impact on the eye-movement responses of the two age groups, they do not display any interaction. Overall, this study suggests that reliance on mechanisms of contextual expectations becomes stronger when we get older.

Keywords: aging; active vision; visual search; contextual expectations; priming; eye-movements.

Introduction

Plates and forks are more likely to be found in kitchens than in bathrooms. Expectations of this sort are constantly generated by our cognitive system to predictively anticipate the state of events, and pro-actively generate appropriate responses (Bar, 2007; Clark, 2013).

Visual attention is a cognitive modality that relies heavily on such contextual expectations to maximize performance in visual tasks (Torralba, Oliva, Castelhano, & Henderson, 2006). In particular, the allocation of visual attention in search tasks improves when semantic information of the target is combined with the contextual information wherein the target is embedded (Malcolm & Henderson, 2010). Moreover, this joint use of target and context information can occur at early stages of search, as long as the quality and quantity of the contextual information available is adequate (Spotorno, Malcolm, & Tatler, 2014).

However, as we age, the visual system deteriorates and performance on visual tasks starts to decrease (Potter, Grealy, Elliott & Andrés, 2012; Brockmole and Logie, 2013). In particular, visual search for a target, manipulated on low-level features (e.g., color, shape and orientation) and embedded in an array of distractors (set of increasing sizes) worsens with increasing age (Humphrey & Kramer, 1997), even though the age-related differences do not manifest to same extent in short visual-working memory (Brockmole, Parra, Sala, & Logie, 2008).

Most of the above studies on aging focused on low-level features, with few other studies approaching this topic from a naturalistic perspective, where knowledge-based rather than stimulus-based processing is examined (Park, 2012). Thus, it remains an open question whether knowledge-based expectancy mechanisms undergo similar changes to those observed on low-level stimulus based information. In this study, we precisely investigate the role of expectations by comparing younger and older observers in a visual search, contextual priming task situated in naturalistic scenes.

Priming paradigms are used to probe, and activate, semantic representations in memory. For example, if a prime word is semantically related to a target word (plant-grass), participants' lexical decision responses are shorter than when the prime word does not belong to the same semantic category (Stanovich & West, 1981).

Effects of priming manifest also during visual attention. Mudrik and Koch (2013), for example, show in a congruency detection task that contextual priming occurs unconsciously, with scenes presented for as little as 33ms. Moreover, repeated exposure to the same visual scene can improve the speed of search, but only for the same target object (Vo & Wolfe, 2012). Additionally, as aging impacts on executive control and inhibition suppression, search performance in seniors would benefit from valid cues, and conversely, take longer to recover from invalid cues (e.g., Müller-Oehring, et. al. 2013)

The main goal of the current study is to establish how contextual expectations are utilized by the visual system by younger (under 25 years) and senior (over 65 years) participants during a search task. In particular, using a priming paradigm, we probe the congruency of contextual information between two scenes (the prime scene and the search scene), sequentially presented (e.g., bathroom-bathroom, versus, kitchen-bathroom), and the consistency of the target object with the search scene (e.g., a toothpaste versus a fork). In line with research showing that seniors have difficulties with inconsistent cues, we expect them to display a significantly worst search performance when the target object is inconsistent with the primed information. Their difficulties with inhibition suppression would imply an increased difficulty to detect a target object when it inconsistently deviates from the congruently primed contextual expectations.

[1]MB is now at Maastricht University, FPN

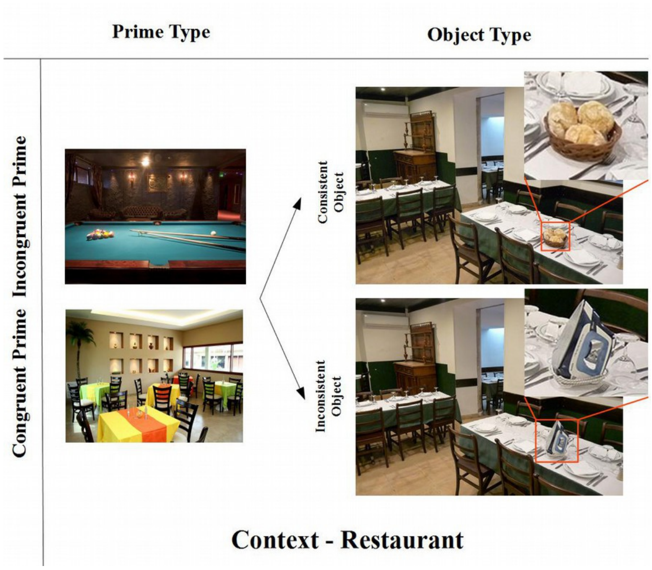

Figure 1: Design. Participants are primed (left column) with either a congruent (restaurant) or an incongruent scene (pool hall), and then see a search scene (right column) wherein a consistent or inconsistent target is placed. The target objects (consistent = 'basket of bread'; inconsistent = 'an iron') are highlighted in red, and zoomed in.

Our study

In this eye-tracking study, we compare the behavioural performance (accuracy and response time), and attentional responses (search latency, first fixation and total gaze) of younger and older participants in a visual search task situated in naturalistic scenes. We investigate how contextual priming and the semantic consistency of the search target mediates performance during aging.

Method

Participants

Fifty-four participants Portuguese native speakers, divided into two age groups: 32 younger (28 women), age 22.63 (SD = 9.03), University students; and 22 older volunteers (19 women), age 67.9 (SD = 4.75), mostly recruited from senior universities, with no medical history of neuropsychological illness Both groups had normal or corrected to normal vision. The study was approved by the Ethics Committee of the Department of Psychology, University of Lisbon (FPUL).

All participants underwent the Montreal Cognitive Assessment Test - Portuguese Version (MoCA) (Simões et al., 2008), used as a criteria to detect potential mild cognitive deficit, especially for the older group. The MoCA was applied on paper with instructions given orally. The young

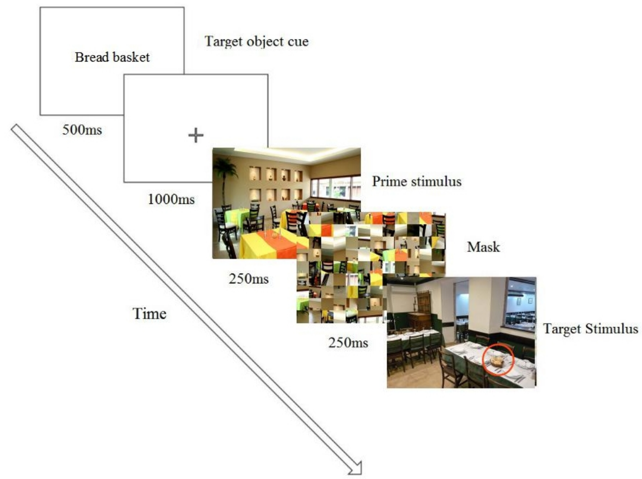

Figure 2: Trial run. Target object and Prime scene are congruent with the search scene (restaurant-bread).

group scored an average of 27.1 out of 30 on the MoCA (SD = 1.83), while the elderly group scored an average of 24 out of 30 (SD = 2.81). The groups significantly differ in their performance, $t(32.8) = 4.46$, $p < 0.001$), however, none of the participants in the elderly group performed below the cut-off score of 19.

Design

Our 2x2x2 experimental design crosses object Consistency (Consistent or Inconsistent), contextual Congruency (Congruent, Incongruent) between the prime scene and search scene, and a between-participants age Group condition (younger or older participant). Refer to Figure 1 for an example of the design material and Figure 2 for the trial run. For example, a participant could be primed with a Restaurant scene, then presented with another Restaurant scene, and previously asked to search for a basket of bread. This example corresponds to the Congruent and Consistent experimental condition.

Material

We utilize 42 experimental scenes, 42 filler scenes and 84 prime scenes, which are all naturalistic, real-world photographs. In particular, the 42 experimental items were drawn from 7 different indoor scenarios (Restaurant, Church, Office, Waiting Room, Bathroom, Living Room, and Kitchen) with 6 items per scenario. The scenes were captured in Lisbon, and surrounding areas, with a congruent or an incongruent object placed by the photographer (refer to section *Validity pre-tests of experimental material* for details on the validation of the consistency between the target objects and the search scene). The position of the target object was counterbalanced with left-right conditions to avoid the development of search strategies. We decided to take photographs of the scenes to obtain a naturalistic visual saliency, and avoid uncontrolled artifacts (e.g., brightness or contrast) that could be introduced when the target object is digitally inserted into the scene. Material was distributed in 4 lists using a Latin-Square Design.

Primes and fillers were selected from various freely available online databases, such as the SUN database (Xiao, Hays, Ehinger, Oliva, & Torralba, 2010) or Flickr. The primes were selected on the basis of their Congruency with respect to the experimental scenes (refer to next section). Moreover, we made sure that the prime scene never contained the target object. The same databases were used for the filler scenes, and we have chosen scenes belonging to similar contexts. No filler trials did included the target object.

Validity pre-tests of experimental material

Firstly, we asked 10 students of the FPUL (6 women) between the ages of 20-25 years (mean = 21.5, SD = 1.433) to name 8 objects they considered typical of each of the 7 contexts. 198 unique objects were produced overall, and among those we selected the 8 most nominated objects for each context (e.g., pillow for context bedroom), as the consistent targets. The 8 inconsistent objects, instead, were drawn from another pool of objects never mentioned by any of the participants. A new group of 20 participants (12 women, mean age = 21.2, SD = 1.23), was then asked to judge the consistency between the target objects previously selected (8 consistent, 8 inconsistent) and the embedding context on a 6-point Likert scale (1 – Totally inconsistent; 6 – Totally consistent). We selected the 6 objects that had highest (consistent) and lowest (inconsistent) scores for each context (e.g., Soap – Bathroom: mean = 5.9, SD = 0.31; Flashlight-Bathroom: mean = 1.3, SD = 0.67). Finally, a new group of 12 participants (6 women, mean age = 22.5, SD = 1.87) was asked to respond to another Likert scale questionnaire and rate: (a) the contextual congruency between the prime scene and the search scene, (b) the typicality of the prime scene and search scene given their context, (c) the consistency of the target object within the search scene, and (d) the ease of recognizing the target object given its linguistic denotation.

The purpose of this last questionnaire was to empirically validate the Congruency and Consistency manipulations. Results show that: (a) the difference between congruous prime scene/search scene (mean = 5.4, SD = 1.09) and their incongruent version (mean = 1.5, SD = 0.92), was statistically significant ($t(489) = 44$, $p < 0.001$); (b) both prime scene (mean = 5, SD = 0.87) and search scene (mean = 5 , SD = 0.86) were considered to be highly representative examples of their context ($t(503) = 0.167$, $p = 0.87$); (c) the target objects consistent with the search scene (mean = 5.1, SD = 1.6) were rated as significantly more likely than their inconsistent counterparts (mean = 1.7, SD = 1.2) $t(470) = 28$, $p < 0.001$); and (d) the target objects were rated as highly identifiable by their linguistic denotation (mean = 5.3 SD = 1.7).

Equipment

Visual stimuli were presented on a LG Flatron L194ws 19-inch LCD screen with a 60 Hz refresh rate, 55cm away from participants' eyes. Eye-movements were recorded using an SMI IVIEW X™ HI-SPEED eye-tracker at a sampling rate of 1250 Hz on a 21” screen (1024 x 768 px. image resolution). Viewing was binocular but only the participant’s dominant eye was tracked (determined by a prior parallax test). Chin rest, and forehead support were used to stabilize the head position, and keep viewing distance constant. Participants logged their answers regarding the presence (or not) of the target object in the scene, using a Logitech Cordless Rumblepad 2 controller. They used 3 buttons for: “Yes”, “No” and to advance to the next trial. The experimental protocol was controlled by ExperimentCenter 3.2 (SMI).

Procedure

Participants were first presented with the target object (a written cue for 500ms). Then, a fixation cross appeared in the center of the screen for 1000ms after which the prime scene is presented for 250ms, followed by a mask[2] of the same scene for 250ms. The mask is utilized to disrupt feedback between lower and higher visual areas and keep the perception of the prime scene on a subliminal level (Enns & Di Lollo, 2000). Finally, the search scene appears and the participants can inspect it until they self-terminate the trial by stating whether they saw (or not) the target object.

Each block contained 42 experimental trials and 42 filler trials, for a total of 84 trials, which were randomly presented to each participant. The target object was present in the experimental trials, and absent in the filler items guaranteeing a balanced 50-50 distribution of yes/no responses, as well as keeping the participants engaged in the search task. 4 practice trials were administered at the beginning of the experimental session to familiarize participants with the task. A 15-points calibration procedure was performed at the beginning of the trial; and repeated every 10 trials (8-points) to ensure accurate tracking. The calibration was accepted when angle threshold was ≈ 0.8/1.2 for x and y respectively. On average, the younger population had 0.61º ± 0.42º deviation error on the x-axis. and 0.72º ± 0.58º on the y-axis. The older population had 0.82º ± 0.60º deviation error on the x-axis and 1.12º ± 0.79º on the y-axis. The experiment was explained using written instructions and the eye-tracking session lasted for approximately twenty to thirty minutes.

Analyses

The performance measures analyzed are: (a) response accuracy, a binomial variable for correct and incorrect responses (1, 0), and on correct trials only, (b) the response time to pass to the next trial (in ms.). From the eye-movement responses to the target object, only for correct trials

[2]The masks were generated by splitting the prime scene into 12 by 9 quadrants, 100x100 pixel squares, and randomly shuffle them.

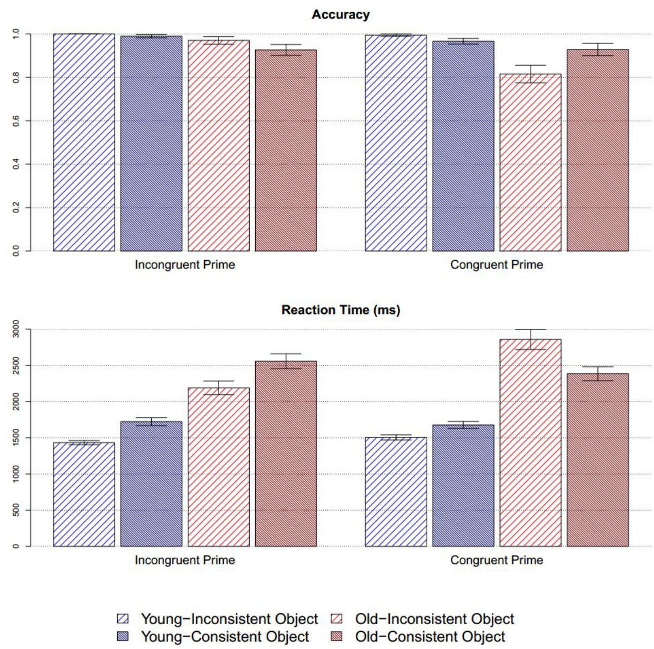

Figure 3: Means and SE for accuracy and reaction time, grouped by Congruency of the prime scene with the target scene (Congruent/Incongruent), Consistency of target object (consistent/inconsistent) and age Group (older/younger). We distinguish between the different experimental conditions using colour and density of lines.

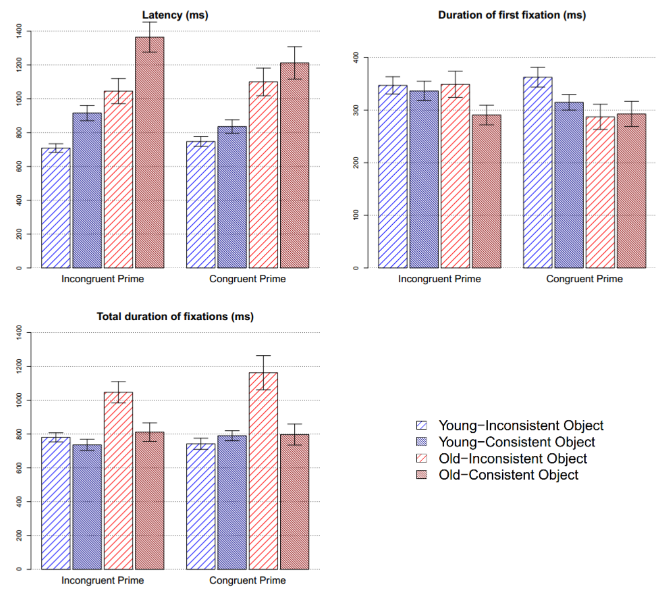

Figure 4: Means (SE) of search latency, first fixation and total duration, grouped by Congruency of the prime scene with the search scene (congruent/incongruent), Consistency of target object (consistent/inconsistent) and age Group (older/younger). We distinguish between the different experimental conditions using color and density of lines.

we compute: (c) the search latency in ms. (i.e., the time from scene onset until the target object is looked at for the first time), (d) the duration of the first fixation in ms., and (e) the total duration of fixations on target-object in ms. These eye-movement measures are commonly used to characterize attentional performance in search tasks (Malcolm & Henderson, 2010). Fixations to the target object AOI were mapped using BeGaze (SMI). Statistical tests on the dependent measures were carried out using mixed-ANOVAs where the between-participant independent variables are **Congruency** between the prime scene and the search scene (2 levels: congruent and incongruent), object **Consistency** (2 levels: consistent and inconsistent) and the within-participant **Group** variable (2 levels: younger and older). All analyses are performed using **R** (ver. 3.0.2, R Core Team, Austria).

Results

Accuracy

In Table 1, we report means and standard deviation for each dependent measure across all experimental conditions, which are also visualized in Figure 3 and 4. We observe a significant main effect of Congruency, whereby prime scenes congruent with search scenes improve the detection of the target object [$F (1, 1153) = 10.2$, $p < 0.01$]. Younger participants are overall more accurate than older participants [$F (1,1153) = 44.4$, $p < 0.001$]. Moreover, we find a significant two-ways interaction Consistency:Group, whereby older participants detect more accurately inconsistent than consistent object [$F (1,1153) = 4.22$ $p < 0.05$] Moreover, we observe a significant three ways interaction Congruency:Consistency:Group, whereby response accuracy of older participants drastically drops when prime scene and search scene are Congruent, but the target object is Inconsistent with the search scene [$F (1,1153) = 14.07$, $p < 0.001$].

Reaction time

Scenes where the target object was consistent with the context of the search scene took longer to be responded [$F (1, 1109) = 17.83$, $p < 0.0001$]. Older participants took longer to respond than younger participants [$F (1, 1109) = 309.82$, $p < 0.0001$]. When looking at the two-ways interaction, we find that a search scene containing a Consistent target object, which was Congruently primed [$F (1, 1109) = 8.98$, $p < 0.01$] yielded faster response times than a scene with an Inconsistent target object; especially in older participants [$F (1, 1109) = 6.77$, $p < 0.01$].

		Incongruent Prime Context		Congruent Prime Context	
		Inconsistent Object	Consistent Object	Inconsistent Object	Consistent Object
		Mean (± SD)	Mean (± SD)	Mean (± SD)	Mean (± SD)
Accuracy	Young	1.0 (± 0)	0.98 (± 0.1)	0.99 (± 0.07)	0.96 (± 0.18)
	Old	0.97 (± 0.17)	0.92 (± 0.26)	0.81 (± 0.39)	0.92 (± 0.26)
Reaction Time (ms)	Young (ms)	1432.7 (± 409.2)	1722.6 (± 745)	1504.8 (± 450)	1676.9 (± 699.8)
	Old (ms)	2189.4 (± 953.2)	2557.6 (± 1065.5)	2858.3 (± 1329.3)	2384.5 (± 871.6)
Latency (ms)	Young (ms)	708.7 (± 374.5)	915.6 (± 612.3)	747.8 (± 378.8)	836 (± 561.6)
	Old (ms)	1045.7 (± 730)	1364.9 (± 890.1)	1100 (± 708.9)	1212.1 (± 838.6)
First Fixation (ms)	Young (ms)	346.9 (± 240.8)	336.4 (± 251.5)	362.5 (± 243.1)	314.7 (± 206.3)
	Old (ms)	348.9 (± 168.3)	290.7 (± 186.3)	287.1 (± 207.4)	292.8 (± 210.1)
Total Fixation (ms)	Young (ms)	780.4 (± 390.4)	735.9 (± 450.3)	742.2 (± 431.2)	789.7 (± 871.8)
	Old (ms)	1046.7 (± 622.3)	811 (± 548.9)	1162.4 (± 871.8)	796.6 (± 545.4)

Table 1: Mean results. Means and standard deviations as a function of the two Prime context conditions and two Object consistency conditions for each of the age groups.

Search latency

Young participants are overall faster than older participants to look at target object for the first time [$F(1, 1109) = 95.81$, $p < 0.0001$]. Search latency is faster when the object is Inconsistent with the search scene, compared to when it is Consistent [$F(1, 1109) = 24.06$, $p < 0.001$]. However, when prime scene and target scene are Congruent, and the target object is Consistent, we observe significantly shorter search latencies [$F(1, 1109) = 4.19$, $p < 0.05$].

First fixation

On the duration of the first fixation, we only find that: (1) consistent target objects are fixated the first time for longer than inconsistent objects [$F(1, 1109) = 4.42$, $p < 0.05$]; and (2) older participants linger on the target object for longer than younger participants [$F(1, 1109) = 5.05$, $p < 0.05$].

Total fixation

On the total fixation duration to the target object, we find main effects of Consistency, whereby participants fixated an Inconsistent target object for longer than when it was Consistent [$F(1, 1109) = 9.10$, $p < 0.01$]; and of Group, with younger participants displaying shorter total duration of fixation than older participants [$F(1, 1109) = 32.96$, $p < 0.001$]. Furthermore, a two-way interaction effect between these two variables was also observed, whereby older participants fixated the target object overall less when it was Consistent with the scene [$F(1, 1109) = 19.92$ $p < 0.001$].

Discussion

Both age groups are clearly influenced by contextual priming, object consistency, as well as by their interaction.

In particular, correct detections of the target object strongly depend on the type of priming participants receive. Incongruent primes improve target detection, but this is more strongly the case when the object is also inconsistent. By probing an unrelated contextual template, participants may become more aware of inconsistency within the search scene. The most interesting result, however, is that the detection performance of older participants abruptly drops when prime and search scenes are congruent, but the target object is inconsistent with the search scene. This result confirms previous literature showing that older adults have more difficulties to inhibit invalid cued information (Müller-Oehring, 2013). Another possibly related explanation is that older participants might rely more strongly on contextual expectations than younger participants. So, when primed with an appropriate context (i.e., restaurant-restaurant), they generate strongly structured expectations for consistent information (e.g., finding the bread on the table of a restaurant). Thus, when the target object violates such expectations, they do not manage to inhibit contextual priming, and the target escapes their attention. These results are largely corroborated by the response times on correct responses, where again we find faster response times with incongruent primes for inconsistent target objects. Crucially, older participants are much slower to respond with inconsistent objects after being primed with contextually congruent information. As just said but from a slightly different angle, this result might indicate that young participants better adapt contextual expectations to the actual search scene, and hence are better able to resolve the mismatch generated by the inconsistency of the target object with the primed expectations. On eye-movement responses, older participants compared to younger participants are overall slower locating the target object and spend more time processing its identity during the first fixation and overall. Moreover, inconsistent objects are found faster than consistent object irrespective of the group, in line with previous literature manipulating this experimental factor (Bonitz & Gordon, 2008). Both groups improve their search latency if the priming is congruent, and the object is consistent. This indicates that appropriate contextual priming can improve performance, but only if the object to be looked at is consistent with the primed information. Crucially, however, we did not find a significant three-ways interaction between Congruency, Consistency and Group; which might indicate that aging does not really mediate the joint influence of Consistency and Congruency on attentional responses.

Conclusion

Expectations are routinely used by the cognitive system to anticipate information about real-word events, and optimize appropriate action-responses (Bar, 2007; Clark, 2013). Visual attention is one of the components that can benefit from contextual expectations in tasks such as visual search, (Malcolm & Henderson, 2010; Spotorno et. al. 2014). Aging, however, might impair attentional performance, and consequently change the way contextual expectations are accessed, and subsequently used, to allocate attention on scenes. For example, visual performance in search tasks, manipulating low-level features, appears to degrade with increasing age (Humphrey & Kramer, 1997). An open challenge was to investigate whether knowledge-based expectations undergo similar degradation, and especially, how attentional mechanisms would be differently used.

In the current study, we used a contextual priming paradigm to probe expectations, and compared visual performance in a search task between younger and older observers. We demonstrated that older observers rely more strongly on consolidated expectations than younger observers. In particular, when the prime and search scene overlapped in context, older participants had a worse performance than the younger, especially if the target object was inconsistent with the primed expectations. This result suggests that older observers might have more difficulties re-computing the semantics of individual objects when primed, and hence compensate for unexpected information. Nevertheless, eye-movement responses did not show a reliable difference in this interaction between congruency and consistency due to aging. Both groups experienced better search performance when the target object was consistent with expectations, but no difference was observed in this interaction due to age.

This study opens intriguing questions on the role of primed contextual expectations in clinical populations affected by neuro-degenerative disease (e.g., Alzheimer); and especially on how would behavioural patterns compare with a normally aging population.

Overall, this work confirms the important role that contextual information plays on the active allocation of visual attention and shows that reliance on it may become stronger as we age.

Acknowledgments

Fundação Amadeu Dias for student award to MB, and Fundação para a Ciência e Tecnologia research fellowship award to MIC (SFRH/BDP/88374/2012), are gratefully acknowledged.

References

Bar, M., 2007. The proactive brain: using analogies and associations to generate predictions. *Trends in cognitive sciences 11 (7),* 280–289.

Bonitz, V. S., & Gordon, R. D. (2008). Attention to smoking related and incongruous objects during scene viewing. *Acta Psychologica, 129*, 255–263.

Brockmole, J., Parra, M., Sala, S., Logie, R. (2008). Do binding deficits account for age-related decline in visual working memory? *Psychonomic Bulletin & Review, 15(3),* 543-547.

Brockmole, J. R., & Logie, R. H. (2013). Age-related change in visual working memory: a study of 55,753 participants aged 8–75. *Frontiers in Psychology, 4.*

Clark, A., 2013. Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences 36 (03),* 310 181–204.

Enns, J. & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences, 4(9),* 345-352.

Humphrey, D. G., & Kramer, A. F. (1997). Age differences in visual search for feature,conjunction, and triple-conjunction targets. *Psychology and Aging, 12,* 704–717.

Malcolm, G. L., & Henderson, J. M. (2010). Combining top-down processes to guide eye movements during real-world scene search. *Journal of Vision, 10(2),* 4.

Mudrik, L., & Koch, C. (2013). Differential processing of invisible congruent and incongruent scenes: A case for unconscious integration. *Journal of Vision, 13(13)*, 24.

Müller-Oehring, E. M., Schulte, T., Rohlfing, T., Pfefferbaum, A., & Sullivan, E. V. (2013). Visual search and the aging brain: discerning the effects of age-related brain volume shrinkage on alertness, feature binding, and attentional control. *Neuropsychology*, *27*(1), 48.

Park, Denise, ed. *Cognitive aging: A primer.* Psychology Press, 2012.

Potter, L. M., Grealy, M. A., Elliott, M. A., & Andrés, P. (2012). Aging and performance on an everyday-based visual search task. *Acta Psychologica, 140(3)*, 208-217.

Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2014). How context information and target information guide the eyes from the first epoch of search in real-world scenes. *Journal of Vision, 14(2)*, 7.

Stanovich, K. E., & West, R. F. (1983). On priming by a sentence context. *Journal of Experimental Psychology: General, 112(1)*, 1.

Torralba, A., Oliva, A., Castelhano, M., & Henderson, J. (2006). Contextual guidance of eye movements and attention in real-world scenes:The role of global features in object search. *Psychological review, 4(113)*, 766–786.

Võ, M. L. H., & Wolfe, J. M. (2012). When does repeated search in scenes involve memory? Looking at versus looking for objects in scenes. *Journal of Experimental Psychology: Human Perception and Performance, 38(1)*, 23.