

Running head: COGNITIVE DEVELOPMENT IN VISUAL SEARCH

**Cognitive development attenuates audiovisual distraction and promotes the
selection of task-relevant perceptual saliency
during visual search on complex scenes**

Clarissa Cavallina¹, Giovanna Puccio¹,

Michele Capurso¹, Andrew J. Bremner^{2,3}, & Valerio Santangelo^{1,4}

1. Department of Philosophy, Social Sciences & Education, University of Perugia, Italy
2. Department of Psychology, Goldsmiths, University of London, UK
3. School of Psychology, University of Birmingham, UK
4. Neuroimaging Laboratory, Santa Lucia Foundation, Rome, Italy

Accepted as an Original Article in Cognition:

Cavallina, C., Puccio, G., Capurso, M., Bremner, A. J., & Santangelo, V. (2018). Cognitive development attenuates audiovisual distraction and promotes the selection of task-relevant perceptual saliency during visual search on complex scenes. *Cognition*, 180, 91-98.

<https://doi.org/10.1016/j.cognition.2018.07.003>

<https://www.sciencedirect.com/science/article/abs/pii/S0010027718301860?via=ihub>

Corresponding Author:

Valerio Santangelo,
Dept. of Philosophy, Social Sciences & Education
University of Perugia
Piazza G. Ermini, 1
06123 Perugia (Italy)
E-mail: valerio.santangelo@unipg.it

ABSTRACT

Searching for a target while avoiding distraction is a core function of selective attention involving both voluntary and reflexive mechanisms. Here, for the first time, we investigated the development of the interplay between voluntary and reflexive mechanisms of selective attention from childhood to early adulthood. We asked 6-, 10-, and 20-year-old participants to search for a target presented in one hemifield of a complex scene, preceded by a task-irrelevant auditory cue on either the target side (valid), the opposite side (invalid), or both sides (neutral). For each scene we computed the number of salient locations (NSL) and the target saliency (TgS). All age groups showed comparable orienting effects (“valid minus neutral” trials), indicating a similar capture of spatial attention by valid cues irrespective of developmental stage. However, only adults demonstrated a suppression of the reorienting effect (“invalid minus neutral” trials), indicating late developments in the reallocation of spatial attention toward a target following auditory distraction. The searching performance of the children (both 6- and 10-year-olds), but not of the adults, was predicted by the NSL, indicating an attraction of processing resources to salient but task-irrelevant locations in childhood; conversely, only adults showed greater performance with increased TgS in valid trials, indicating late development in the use of task-related saliency. These findings highlight qualitatively different mechanisms of selective attention operating at different ages, demonstrating important developmental changes in the interplay between voluntary and reflexive mechanisms of selective attention during visual search in complex scenes.

Keywords: Visual search, complex scenes, crossmodal spatial attention, perceptual saliency, cognitive development.

INTRODUCTION

Selective attention, the ability to search for a target while ignoring distracting stimuli, is a core function of human cognition. In the last decades, a significant body of research has been devoted to understanding the role of selective attention in visual search (Eimer, 2014). There is now a wide consensus that selective attention can affect visual search either voluntarily, i.e., when attention is driven by visual search goals, or involuntarily (reflexively), i.e., when attention is driven by the physical saliency of the stimuli (e.g., Anderson, 2013; see also Klein & Shore, 2000). During search, there is a tight interplay between these two components of selective attention. In fact, voluntary attention devoted to search for a current target (or targets) is continuously challenged by distractors that can involuntarily capture available processing resources.

There is a clear consensus that involuntary aspects of selective attention are relatively stable across childhood and adulthood with voluntary aspects following a U-shaped pattern of development (improvements to young adulthood and decline into old age; e.g., Enns, Brodeur, & Trick, 1998; Rueda et al., 2004; Waszak, Li, & Hommel, 2010). In the domain of visual search, this issue has been typically assessed by comparing search for conjunctions of visual features vs. single features. While the former has been thought to provide an index of top-down attentional control, the latter has been considered to be mediated primarily by bottom-up involuntary orienting. For instance, Donnelly, Cave, Greenway, Hadwin, & Sonuga-Barke (2007) compared feature conjunction visual search performance between child and young adult participants. When searching for conjunctions of features the children searched more slowly and took longer to reject trials than adults did when no target was present, indicating an age-related development of top-down control in visual search. By contrast however, searching for single visual features shows very little, if any development between childhood and adulthood (Trick & Enns, 1998). This pattern of findings is consistent with the general notion that voluntary visual search performance improves with age as a consequence of increasing attentional control across development (for reviews see Atkinson & Braddick, 2012; Scerif, 2010). While this important literature has clarified the distinct

developmental trajectories of voluntary and involuntary attention, the development of the interaction of these two modes of attentional selection during visual search has been largely neglected to date.

One means of examining the interplay between voluntary and involuntary selective attention during visual search is by manipulating the deployment of visual spatial attention through a cuing procedure. In adults, this has typically been assessed by presenting either visual (Briand & Klein, 1987; Carrasco & Yeshurun, 1998; Cohen & Ivry, 1989; Muller & Rabbitt, 1989; Prinzmetal, Presti & Posner, 1986) or auditory (Perrott, Saberi, Brown & Strybel, 1990; Rudmann & Strybel, 1999; Vu, Strybel & Proctor, 2006) spatial cues before the presentation of the visual search scene. For instance, in their seminal study Prinzmetal and colleagues (1986) asked participants to search for a colored letter among distractor letters. On most trials, the participants were pre-cued by a visual peripheral signal to the spatial location where the target letter was presented (valid trial), while on the remaining trials a non-target location was cued (invalid trial). Prinzmetal and colleagues found more accuracy in valid than invalid trials. Although clearly demonstrating an effect of selective spatial attention on visual search, this study failed to demonstrate whether the effect was driven by: i) a facilitation of visual search by selective attention, ii) distraction by task-irrelevant stimuli, or iii) both. Key to disentangling these interpretations, is the deployment of a spatially “neutral” pre-cue condition which permits the assessment of both the orienting component (via a comparison of valid and neutral trials) and the reorienting component (via a comparison of invalid and neutral trials).

The current study therefore aimed to investigate developmental changes in the deployment of selective attention during visual search, from childhood to adulthood. Moreover, this developmental design also yields insight into the mechanisms of attention in general (i.e., in the mature adult as well as other age groups; see Enns et al., 1998; Matusz et al., 2015) by demonstrating how the interplay between voluntary and reflexive components of selective attention is affected by developmental changes in the control of selective attention. Thus, we presented 6-, and 10-year-old children and young adults (20-year-olds) with visual scenes. Each scene included several objects,

but only one agentive element (an animal or a human) which was the target stimulus, located on the right or left side of the scene. The participants were asked to search for the target (i.e., a voluntary attention task). Each scene was preceded by a sound (i.e., a burst of white noise serving as an auditory¹ pre-cue; duration = 50 ms) that was presented 50 ms before the scene. The auditory pre-cue was equiprobably presented either on the target side (valid trials), on the opposite side (invalid trials), or on both sides (neutral trials). Consequently, the auditory pre-cue was entirely task-irrelevant, that is, it was not a useful strategy to voluntarily pay attention to the auditory cue in order to carry out the visual search task. This cuing procedure therefore enabled us to disentangle the specific contributions of the orienting (indexed by valid minus neutral trials) and reorienting (invalid minus neutral trials) components of reflexive crossmodal attentional mechanisms during the voluntary searching task.

Studies of visual search typically use very simple, repetitive and unrealistic displays. However, there is now a growing consensus among researchers (and increasing efforts are being made in this respect) that it is necessary to use more naturalistic stimuli and behavioral tasks in order to gain a valid picture of how spatial attention works in ecological situations (for reviews see Felsen & Dan, 2005; Peelen & Kastner, 2014). For this reason, the visual searches we asked participants to make were to complex and unrepeated cartoon-like scenes. The use of complex scenes also allowed us to characterize the impact of visual features included in each scene, thus to measure also reflexive orienting towards task-relevant or task-irrelevant visual elements. To do this, we computed a “saliency map” for each scene using a well-known algorithm (Itti, Koch & Niebur, 1998) based on local discontinuities in line orientation, intensity contrast, and color opponency (see Fig. 1B). Saliency maps are topographic representations of dominant locations in a visual scene that are more likely to be attended to and processed during scene exploration (see, for a review, Itti & Koch, 2001). Each dominant (or salient) location in the scene is represented by a saliency “peak”, that is, the highest saliency value within that location. The higher the value of the location’s

saliency peak, the more likely it is that that location will be attended to and processed during scene exploration.

By means of the above-described saliency maps, we computed two saliency-related indexes for each scene. The first one was based on the computation of the “number of salient locations” (NSL) within each scene. We used this index to characterize the dispersion of attentional resources over locations that are not necessarily involved with the detection of the to-be-searched target. If searching performance decreased (i.e., if we observed longer reaction times) as a function of an increased number of salient locations, this would mean that participants waste spatial attentional resources on salient but task-unrelated locations. In this sense, this index is a measure of an inefficient search strategy. The second saliency-related index was based on the level of perceptual saliency related to the current target stimulus, namely the “target saliency” (TgS) index, which corresponded to the saliency peak falling inside the to-be-searched target in the scene. If searching performance increased (i.e., if we observed shorter reaction times) as a function of increased target saliency, this would mean that participants could successfully use task-relevant (i.e., target-related) sensory information to accomplish the searching task.

If age-related level of attentional control plays a critical role on the interplay between voluntary and reflexive mechanisms of selective attention, we would expect larger reflexive orienting and reorienting effects in children as compared to young adults, resulting from a lesser capacity to remain focussed on task-relevant target features while filtering out task-irrelevant information (Enns et al., 1998; Rueda et al., 2004; Waszack et al., 2010). Considering visual saliency, we expected that the different age groups would provide opposite effects in relation to the two saliency indexes. Specifically, we expected that the number of salient locations would predict visual search performance in children but not in adults, indicating a “dispersion” of attentional resources over the different salient locations irrespective of their task-relevancy. By contrast, we expected that target saliency would predict visual search performance in adults but not in children,

demonstrating a greater capability to select task-relevant information based on low-level perceptual saliency.

METHODS AND MATERIALS

Participants

One hundred and nine healthy participants volunteered for and took part the study. The sample included: forty children attending the first grade of the “G. Pierluigi” primary school in Perugia, Italy, thirty-eight children attending the fifth grade of the same school, and thirty-one university students (participant characteristics for each group are presented in Table 1). The sample size was estimated with G*Power 3.1.9.2 taking into account: predicted effect size (0.25), power (95%), and significance level (0.05; ANOVA, repeated measures, within-between interaction). This indicated a minimum sample size of 30 participants per group. All of the participants had normal or corrected-to-normal vision and were naïve to the main purpose of the study, which was conducted in accordance with the research ethics principles of the Declaration of Helsinki. All of the adult participants provided informed consent. Parental consent was obtained for each child participant.

Stimuli and task

Cartoon scenes served as visual scenes for the visual search task (see Fig. 1A). These were created using GoAnimate (GoAnimate © 2016; <https://goanimate.com/>) at a resolution of 1024x768 pixels, and involved either internal (a kitchen, a bathroom etc.) or external scenes (a garden, a street etc.). Each scene included several objects, but only one agentive element which was either an animal (a dog, a cat, a rabbit, etc.) or a human (a young man, a young woman, etc.). These agentive elements were the search targets in the scenes.

The participants sat in front of a laptop computer placed with the screen at approximately 50 cm from the viewer (picture size = 29 x 22 degrees of visual angle). Two loudspeakers were placed one on each side of the computer monitor. The visual search task comprised a sequence of eight

blocks of trials. Each block of trials began with the presentation of a display of the target which was to be searched for in the following visual scenes (see Fig. 1A). This target display was presented until the participant was satisfied that he/she could identify the target and wished to continue. When ready to start the search task, the participants pressed either of the two response keys (“D” or “L”). After a gap of 2000 ms in which a blank screen was presented, the first visual scene was presented. Each block included nine such visual scenes. Each of the scenes was displayed for 3000 ms, followed by an inter-stimulus interval (ITI) of 2000 ms, in which a blank screen was presented. The participants’ task was to localize the target in the each scene and to press as quickly and accurately as possible one of the two response keys according to the target location: they had to press “D” on the laptop keyboard when the target was on the left side of the scene (with respect to the central vertical meridian), and to press “L” when the target was on the right side of the scene. After all nine visual scenes had been displayed a new block of trials began, starting with the target display.

Each scene was preceded by the presentation of burst of white noise on either the left, right of both sides of the screen. The sound had a duration of 50 ms and was presented 50 ms before the scene. The sound was equiprobably presented on either the left or right of the scene, and the children were informed that this noise was not informative about the target location. In “valid” trials the sound was presented on the same side as the target in the scene, while in “invalid” trials the sound was presented on the opposite side as the target; finally, in “neutral” trials the sound was presented simultaneously on both sides (loudspeakers) of the scene.

The experiment included 8 blocks, each containing 9 trials, of which 3 were valid, 3 neutral, and 3 invalid. This resulted in a total of 72 trials, 24 for each cue type. The order of blocks and the order of scenes and conditions within each block was randomized across participants. Each scene was randomly assigned to the valid, neutral or invalid conditions, with the constraint of having 3 valid, 3 neutral and 3 invalid scenes within each block. Before the experimental session participants practiced with a short training session comprising 4 blocks of 6 trials each (2 valid, 2 neutral and 2 invalid) to familiarize with the task.

Computation of low-level sensory saliency indexes

To investigate whether low-level sensory features (i.e., perceptual saliency) predicted visual search performance in the three groups of participants we analyzed each visual scene with the Saliency Toolbox 2.2 (<http://www.saliencytoolbox.net/>). This allowed us to compute saliency maps using local discontinuities in line orientation, intensity contrast, and color opponency (Itti et al., 1998; see Fig. 1B). Saliency maps allowed us to extract from each scene two different indexes related to low-level sensory features. The first index was related to the saliency level of each target in the scene, namely, the “target saliency” index (TgS), defined as the value of the saliency peak (i.e., the highest saliency value) falling inside the target shape (e.g., the yellow puppy in Fig. 1B). The second index was instead related to the quantity of saliency peaks included within each scene, namely, the “number of salient locations” index (NSL; note that this index was computed with the target included in the scene). To rule out the possibility that saliency-related effects were explicable by other confounding factors we also computed two other indexes for each scene, namely the “target size” and “horizontal target eccentricity” indexes. Previous studies have shown that these indexes affect visual search performance, with bigger targets and more eccentric targets (i.e., those close to the display center) being easier to be detected (see, e.g., Brand, Oriet, Johnson & Wolfe, 2014; Gruber, Muri, Mosimann, Bieri, Aeschmann, Zito, Urwyler, Nyffeler & Nef, 2014). These indexes were therefore included in the regression model described below. Target size was defined as the area occupied by the target. Using CorelDraw Graphics Suite v. 12 (Copyright © 2016 Corel Corporation) we manually drew a mask around each target. We then computed the proportion of pixels included within the target mask with respect the overall number of pixels of the picture. Horizontal target eccentricity was defined as the distance (converted in degrees of visual angle) between the vertical meridian of the scene and the horizontal coordinate of the centre of mass of the target. The latter was computed by averaging horizontal and vertical coordinates of each pixel

belonging to the target object (see, e.g., Santangelo, Di Francesco, Mastroberardino & Macaluso, 2015).

Data analysis

The participants' reaction time and accuracy data were analyzed with SPSS 21 (Statistical Package for Social Science) using a hierarchical approach. First, we performed mixed analyses of variance (ANOVAs) on the reaction time and accuracy data, with the between-participants factor of Age group (6, 10, or 20 y.o.) and the within-participants factor of Cue type (valid, neutral, invalid), to assess whether searching performance varied across cue type conditions and age groups. In order to investigate significant interactions of Age group with Cue type, we followed these up with one-way ANOVAs which separately investigated effects of age group (6, 10, or 20 y.o.) on orienting effects (valid-neutral trials) and reorienting (invalid-neutral trials) effects. These one-way ANOVAs directly examine our hypotheses concerning age-related decreases in orienting and reorienting in selective attention from childhood to adulthood. The Huynh-Feldt procedure was used to correct for any violations of sphericity.

To investigate the impact of low-level perceptual saliency on visual search performance we used a multivariate multiple regression (MMR) model. MMR is a powerful tool for modelling the linear relationship between more than one independent variable (or predictor) and more than one dependent variable, all at once. The current MMR model included TgS and NSL as predictors. Target size and horizontal target eccentricity were also included in the model in order to rule out any potentially confounding explanations of visual saliency effects. While this allowed us to control the effects of possible confounds whilst assessing the effect of TgS and NSL on search performance, the analysis of the impact of target size and horizontal target eccentricity on subjects' performance falls outside the aims of the current study. Mean RT under valid, neutral and invalid cue conditions in the three age group were entered as dependent variables. The approach of using multiple predictors in MMR – instead of carrying out separate regression models – has the

advantage of estimating the particular influence of each predictor while simultaneously controlling for the influence of the other predictors and potentially confounding factors. Before these analyses we verified that our data did not violate the assumption of homoscedasticity.

RESULTS

RT and accuracy data are summarized in Figure 1C. The ANOVA on the RT data revealed a main effect of Age group [$F(1, 106) = 58.0, p < .001, \eta^2 = .522$]. Newman-Keuls post-hoc comparisons indicated faster responses for the 20 y.o. (469 ms) than for 10 y.o. participants (575 ms; $p = .002$), who in turn made faster responses than the 6 y.o. participants (822 ms; $p < .001$). The RT analysis also revealed a main effect of Cue type [$F(2, 212) = 30.3, p < .001, \eta^2 = .222$]. Post-hoc comparisons showed that the participants' responses were faster for valid (605 ms) than for neutral trials (621 ms; $p < .001$); in turn, responses for neutral trials were faster than for invalid trials (640 ms; $p < .001$). Most importantly, the RT analysis revealed a significant Age group x Cue type interaction [$F(4, 212) = 2.9, p = .024, \eta^2 = .051$], indicating a different effect of Cue type across the three Age groups. This was further investigated by two one-way ANOVAs, one comparing the orienting (valid-neutral trials) and the other the reorienting (invalid-neutral trials) effect across the three age groups' orienting effects (see Tab. 1). The analysis of the orienting effect did not reveal any significant difference [$F(2, 106) < 1, n.s.$], while the magnitude of reorienting effects significantly varied across the three age groups [$F(2, 106) = 3.3, p = .042, \eta^2 = .058$]. Post-hoc comparisons revealed that the magnitude of reorienting for adults (5 ms) was smaller than for 6 y.o. children (33 ms; $p = .031$), who, in turn, did not differ from the 10 y.o. children (18 ms; $p = .161$).²

The 3x3 mixed ANOVA on the accuracy data revealed a pattern of results which was consistent with that gleaned from the RT data. There was a main effect of Age group [$F(1, 106) = 6.5, p = .002, \eta^2 = .109$], indicating that the 20 y.o. (99.0%) and the 10 y.o. groups (98.5%; not differing between them, $p = .329$) were more accurate than the 6 y.o. group (97.1%; $p = .002$ and $p = .015$, respectively). There was also a main effect of Cue type [$F(2, 212) = 11.4, p < .001, \eta^2 =$

.097], indicating lower accuracy following invalid (97.2%) than following neutral cues (98.5%; $p < .001$), which did not differ from valid cues (98.9%; $p = .247$). There was also an Age group x Cue type interaction [$F(4, 212) = 2.8, p = .035, \eta^2 = .050$]. As before, we compared the magnitude of orienting and reorienting effects across the different age groups. The analysis of the orienting effect did not revealed any significant difference [$F(2, 106) < 1, n.s.$], while the magnitude of reorienting effects significantly varied across the three age groups [$F(2, 106) = 3.1, p = .049, \eta^2 = .055$]. Post-hoc comparisons revealed that the magnitude of reorienting for adults (0.3%) was smaller than for 6 y.o. children (-2.2%; $p = .026$), which, in turn, did not differed from 10 y.o. children (-1.9%; $p = .755$).³

Finally, we assessed whether and at which age low-level saliency indexes (i.e., TgS and NSL) related to visual scenes predicted search performance following valid, neutral or invalid auditory cues using the MMR model. Figure 1D summarizes the standardized regression coefficients (betas) related to the two predictors, NSL (left graph) and TgS (right graph). While positive beta values indicated positive covariations between the predictor variables (i.e., NSL or TgS) and the dependent variables (i.e., the reaction times in the different cuing conditions and age groups), negative beta values indicated negative covariations. For the TgS predictor betas were negative as expected, indicating that the higher the target saliency the faster the searching performance. However, the regression was statistically significant only following auditory valid cues in the 20 y.o. group (see the asterisk in Fig. 1D, graph on the left, and Tab. 2 for statistical details), while the level of TgS did not modulate searching performance of younger participants. Conversely, the NSL predictors showed, as expected, a positive relation with searching performance, indicating that the greater the number of salient locations in the scene, the poorer the searching performance. However, this negative impact of NSL on searching performance was statistically significant only in the in the 6 and 10 y.o. groups following valid and invalid auditory cues, and not in the adult group.

DISCUSSION

The main aim of the current study was to investigate the development across childhood to adulthood of the interplay between voluntary (i.e., searching for the target) and reflexive (i.e., involuntary capture by auditory pre-cues or low-level visual features) mechanisms of selective attention during visual search. To address this aim, we presented young adults, 6- and 10-year-old participants with complex (cartoon-like) scenes for a to-be-searched target. Each scene was preceded by an uninformative sound that was either spatially-congruent (valid), or -incongruent (invalid) with the target side, or spatially-neutral. Our results showed that searching performance was strictly dependent on the participants' ages. We found faster searching performance for young adults than for 10-year-olds, who in turn performed better than the 6-year-olds. Although slower than adults, 10-year-old children showed overall a similar level of accuracy than adults (but see the reorienting effects discussed below). Taken together these findings highlight the expected gradual development in attentional control needed to efficiently perform the current visual search task, in line with extant literature on cognitive development of attention-related mechanisms (Atkinson & Braddick, 2012; Scerif, 2010).

Along with these general effects, our use of peripheral auditory pre-cues allowed us to selectively highlight age-related differences related to the orienting and reorienting of selective attention during visual search. Contrary to our prediction, we found comparable orienting effects (valid minus neutral trials) across the three groups of subjects, indicating that the efficacy with which peripheral and uninformative sounds capture visuospatial attention over the left or right visual hemifield (e.g., Spence & Driver, 1997) does not develop significantly beyond 6 years of age. This finding sheds new lights on the classical debate concerning whether so-called "reflexive" attention is actually automatic. Some evidence indicates that exogenous spatial orienting is not truly automatic, at least when assessed crossmodally (see, for a review, Santangelo & Spence, 2008). For instance, Santangelo, Olivetti Belardinelli and Spence (2007) found that focusing voluntary attention on a highly perceptual demanding task such as a rapid serial visual search suppressed orienting effects towards peripheral visual or auditory spatial cues. Similarly reduction in activity in

visual cortex during processing of peripheral stimuli is seen when participants are undertaking central demanding perceptual tasks (see Schwartz, Vuilleumier, Hutton, Maravita, Dolan, & Driver, 2005). However, later studies highlighted that orienting effects can still be elicited by peripheral spatial cues, as long as the stimulus onset asynchrony between the spatial cue and target was short enough (80 ms) not to be interrupted by the letter presentation of the rapid serial visual presentation task, determining a central reorienting of spatial attention (see Santangelo, Botta, Lupianez, & Spence, 2011). These latter findings demonstrate that task-irrelevant abrupt onsets cannot be entirely overridden by top-down attentional control, although attentional capture effects are dramatically reduced by an ongoing perceptually demanding task (see Theeuwes, 2010, for a review of consistent findings). The current findings further extend the notion that peripheral cues capture spatial attention resources even under concurrent conditions requiring voluntary attention, i.e., the searching task. Crucially, here we show for the first time that the automaticity of crossmodal orienting effects of peripheral cues is not related to the level of attentional control (age group), at least between childhood (6 y.o. children) and adulthood.

We found instead substantial differences between children and adults when looking at the reorienting effects (invalid minus neutral trials). While children showed comparable reorienting effects, in adults we observed no significant reorienting effect in terms of both reaction time and accuracy (only 5 ms and 0.3% in absolute terms, respectively). This indicates that adults, after being distracted by the invalid sound, were faster to reallocate their spatial attention towards the target side (also ignoring low-level saliency, see discussion below). Our interpretation of this is that the higher level of attentional control in adults as compared to children contributes to an improved control of voluntary attention towards task-relevant target features (Atkinson & Braddick, 2012; Scerif, 2010). Commensurate with this interpretation, we also observed decreases in search accuracy following specifically invalid auditory cues in both the 6 and 10 y.o. groups.

The current study also investigated developmental differences in the impact of low-level perceptual saliency on attentional orienting/reorienting mechanisms during visual search. Low-level

saliency has been shown to successfully predict overt orienting of spatial attention (i.e., it predicts fixation patterns during free viewing of complex scenes, for both pictures (Borji, Sihite, Itti, 2013; Einhäuser, Spain, & Perona, 2008; Elazary & Itti, 2008; Santangelo et al., 2015), and videos (Carmi & Itti, 2006; Nardo, Santangelo, & Macaluso, 2011, 2014; see, for a review, Itti & Koch, 2000). Highly salient perceptual stimuli, which grab spatial attention, have also been shown to influence post-perceptual processing (by gaining access to memory representation, Fine & Minnery, 2009; Melcher and Piazza, 2011; Pedale & Santangelo, 2015; Santangelo & Macaluso, 2013; see, for a review, Santangelo, 2015). However, the role of perceptual saliency in predicting fixation patterns significantly decreases during visual search tasks (Einhäuser, Rutishauser, & Koch, 2008; Henderson, Brockmole, Castelhamo, & Mack, 2006; Underwood, Foulsham, van Loon, Humphreys, & Bloyce, 2006). This suggests a strict influence of attentional control on the way an observer explore a complex visual scene when attempting to carry out a specific task. Interestingly, if attentional control is constrained, perceptual saliency is a significant predictor of overt exploration, as illustrated by recent investigations of neurologically damaged patients affected by visual agnosia (Foulsham, Barton, Kingstone, Dewhurst, & Underwood, 2009; see also Mannan, Kennard, & Husain, 2009). These patients are severely impaired at recognizing objects or scenes, and therefore diverge from healthy volunteers in their understanding (i.e., attentional control) of the scene. Consistently, during visual search fixated locations generally revealed a closer overlap with low-level saliency maps in agnosia patients than in control subjects.

In agreement with this literature, we showed here that saliency-related effects are much more readily apparent at earlier stages of cognitive development. As predicted, we showed different effects of the specific saliency indexes on search performance at different ages. The number of salient locations (NSL) predicted searching performance in children following valid or invalid auditory cues, but not so in adults. We interpreted this finding as indicating a “dispersion” of attentional resources over the different salient locations irrespective of their task-relevancy: In fact, the greater the number of salient locations, the slower children’s RT, in agreement with the notion

of a serial search across the different salient locations in the scene (Treisman & Gelade, 1980). This interpretation is compatible with recent evidence showing that children are less focussed than adults on task-relevant features in attention and in a wider set of exploratory behaviours (e.g., Bonawitz, Shafto, Gweon, Goodman, Spelke, & Schulz, 2011; Bremner, Mareschal, Destrebecqz & Cleeremans, 2007; Fisher, Godwin & Seltman, 2014; Plebanek & Sloutsky, 2017).

This dispersion index (NSL) did not predict adults' searching performance. Rather, adults were found to process and use low-level perceptual information which was specifically related to the current target, with faster RTs as a function of increased target saliency (TgS), but only following valid auditory cues. This means that the orienting effect of spatial attention observed in the adult groups is qualitatively different from the orienting effects observed in the children groups. Adults processed target saliency and used this bottom-up information as an efficient target search strategy, starting their search with the most salient peaks of the scene. So for adults, the more the target corresponded to the greater saliency peak, the faster was their target search. By contrast, both 6 and 10 y.o. children failed to show any selective advantage for highly salient targets, corroborating the notion that they dispersed their spatial attentional resources over each single saliency peak in the scene irrespective of its saliency level (i.e., the NSL effect previously discussed).

Finally, our analyses revealed that saliency-related factors were never good predictors of searching performance when visual scenes were preceded by neutral auditory cues, across the three groups of participants. This finding points out that there is a strict relationship between the deployment of selective attention and the possibility to process low-level sensory features (i.e., the perceptual saliency of the scene). While the presentation of a peripheral auditory cue (either valid or invalid) triggered the processing of saliency-related information during target searching, the simultaneous presentation of a bilateral sound (i.e., a neutral spatial cue perceived as central) may have led to internally-driven search. We speculate that visual searching paired with neutral auditory cues might be more subject to volitional control, thus bypassing the impact of low-level sensory

features on searching performance (e.g., Einhäuser et al., 2008; Henderson et al., 2006; Underwood et al., 2006). A similar increase in volitional control might be assumed after invalid cues in the adult group, suppressing the processing of target-related saliency.

To conclude, we have reported here the results of a crossmodal visual search task performed on complex scenes by young adults, 6-, and 10-year-old children. This design highlighted qualitatively different spatial attentional mechanisms related to age. Notwithstanding general performance differences (RTs and accuracy) indicating a gradual development in searching tasks, both 6- and 10-year-old children showed similar attentional mechanisms, with comparable orienting and reorienting effects towards valid and invalid auditory spatial cues. Children's search, triggered by valid or invalid auditory cues, made use of any salient locations in the scene. This rather inefficient search strategy contrasts with that used by young adults who made use of low-level target-related perceptual saliency during search. Their search was triggered by valid auditory cues, with any reorienting seemingly suppressed by their greater attentional control. Overall, these findings demonstrate the importance of testing the interplay between voluntary and reflexive mechanism of attentional orienting at variable at different ages: Here we highlighted developmental changes in the dynamic interaction of these orienting mechanisms in controlling the distribution of attention resources during search on complex visual scenes.

Acknowledgements

This study was supported by a Basic Research Grant (“Fondo Ricerca di Base 2015”) awarded by the University of Perugia to V. S. and M. C.

References

- Atkinson, J., & Braddick, O. (2012). Visual attention in the first years: typical development and developmental disorders. *Developmental Medicine and Child Neurology*, *54*, 589-595.
- Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*, 322-330.
- Borji, A., Sihite, D. N., & Itti L. (2013). Objects do not predict fixations better than early saliency: A re-analysis of Einhauser et al.'s data. *Journal of Vision*, *13*, 18.
- Brand, J., Oriet, C., Johnson, A. P., & Wolfe, J. M. (2014). Flexible cue combination in the guidance of attention in visual search. *Acta Psychologica*, *153*, 129-138.
- Bremner, A. J., Mareschal, D., Destrebecqz, A., & Cleeremans, A. (2007). Cognitive control of sequential knowledge in 2-year-olds: Evidence from an incidental sequence-learning and-generation task. *Psychological Science*, *18*, 261-266.
- Briand, K. A., & Klein, R. M. (1987). Is Posner's "beam" the same as Treisman's "glue"?: On the relation between visual orienting and feature integration theory. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 228-241.
- Carmi, R., & Itti, L. (2006). Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Research*, *46*, 4333-4345.
- Carrasco, M., & Yeshurun, Y. (1998). The contribution of covert attention to the set-size and eccentricity effects in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 673-692.
- Cohen, A., & Ivry, R. (1989). Illusory conjunctions inside and outside the focus of attention. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 650-663.
- Dalton, P., & Spence, C. (2007). Attentional capture in serial audiovisual search tasks. *Perception & Psychophysics*, *69*, 422-438.

- Donnelly, N., Cave, K., Greenway, J. A., Hadwin, J. S., & Sonuga-Barke, E. (2007). Visual search in children and adults: Top-down and bottom-up mechanisms. *Quarterly Journal of Experimental Psychology*, *60*, 120-136.
- Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision*, *8*, 18.1-26.
- Einhäuser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, *8*, 1-19.
- Elazary, L., & Itti, L. (2008). Interesting objects are visually salient. *Journal of Vision*, *8*, 3.1-15.
- Enns, J. T., Brodeur, D. A., & Trick, L. (1998). Selective attention over the lifespan: Behavioral measures. In J. E. Richards (Ed.), *Cognitive neuroscience of attention: A developmental perspective*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Felsen, G., & Dan, Y. (2005). A natural approach to studying vision. *Nature Neuroscience*, *8*, 1643-1646.
- Fine, M. S., & Minnery, B. S. (2009). Visual salience affects performance in a working memory task. *Journal of Neuroscience*, *29*, 8016-8021.
- Fisher, A. V., Godwin, K. E., & Seltman, H. (2014). Visual environment, attention allocation, and learning in young children: When too much of a good thing may be bad. *Psychological Science*, *25*, 1362-1370.
- Foulsham, T., Barton, J. J., Kingstone, A., Dewhurst, R., & Underwood, G. (2009). Fixation and saliency during search of natural scenes: the case of visual agnosia. *Neuropsychologia*, *47*, 1994-2003.
- Gruber, N., Müri, R. M., Mosimann, U. P., Bieri, R., Aeschmann, A., Zito, G. A., Urwyler, P., Nyffeler, T., & Nef T. (2014). Effects of age and eccentricity on visual target detection. *Frontiers in Aging Neuroscience*, *5*, 101.
- Henderson, J. M., Brockmole, J. R., Castelano, M. S., & Mack, M. (2006). Visual saliency does not account for eye-movements during visual search in real-world scenes. In: R. van Gompel,

- M. Fischer, W. Murray, & R. Hill (Eds.), *Eye movement research: Insight into mind and brain* (pp. 537-62). Oxford: Elsevier.
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2, 194-203.
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 20, 1254-1259.
- Klein, R. M., & Shore, D. (2000). Relations among modes of visual orienting. In S. Monsell & J. Driver (Eds.), *Attention and performance XVIII: Control of cognitive processes* (pp. 195-208). Cambridge, MA: MIT Press.
- Lavie, N. (2005). Distracted and confused? Selective attention under load. *Trends in Cognitive Sciences*, 9, 75-82.
- Mannan, S. K., Kennard, C., & Husain, M. (2009). The role of visual salience in directing eye movements in visual object agnosia. *Current Biology*, 19, R247-248.
- Matusz, P. J., Broadbent, H., Ferrari, J., Forrest, B., Merkle, R., & Scerif, G. (2015). Multi-modal distraction: insights from children's limited attention. *Cognition*, 136, 156-165.
- Melcher, D., & Piazza, M. (2011). The role of attentional priority and saliency in determining capacity limits in enumeration and visual working memory. *PLoS One*, 6, e29296.
- Müller, H. J., & Rabbitt, P. M. (1989). Spatial cueing and the relation between the accuracy of "where" and "what" decisions in visual search. *Quarterly Journal of Experimental Psychology A*, 41, 747-773.
- Nardo, D., Santangelo, V., & Macaluso, E. (2011). Stimulus-driven orienting of visuo-spatial attention in complex dynamic environments. *Neuron*, 69, 1015-1028.
- Nardo, D., Santangelo, V., & Macaluso, E. (2014). Spatial orienting in complex audiovisual environments. *Human Brain Mapping*, 35, 1597-1614.

- Pedale, T., & Santangelo, V. (2015). Perceptual salience affects the contents of working memory during free-recollection of objects from natural scenes. *Frontiers in Human Neurosciences*, 9, 60:1-8.
- Peelen, M. V., & Kastner, S. (2014). Attention in the real world: toward understanding its neural basis. *Trends in Cognitive Sciences*, 18, 242-250.
- Perrott, D. R., Saberi, K., Brown, K., & Strybel, T. Z. (1990). Auditory psychomotor coordination and visual search performance. *Perception & Psychophysics*, 48, 214-226.
- Plebanek, D. J., & Sloutsky, V. M. (2017). Costs of selective attention: When children notice what adults miss. *Psychological Science*, 28, 723-732.
- Prinzmetal, W., Presti, D. E., & Posner, M. I. (1986). Does attention affect visual feature integration? *Journal of Experimental Psychology: Human Perception and Performance*, 12, 361-369.
- Rudmann, D. S., & Strybel, T. Z. (1999). Auditory spatial facilitation of visual search performance: effect of cue precision and distractor density. *Human Factors*, 41, 146-160.
- Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., & Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia*, 42, 1029-40.
- Santangelo, V. (2015). Forced to remember: When memory is biased by salient information. *Behavioural Brain Research*, 283, 1-10.
- Santangelo, V., Botta, F., Lupianez, J., & Spence, C. (2011). The time-course of attentional capture under dual-task conditions. *Attention, Perception & Psychophysics*, 73, 15-23.
- Santangelo, V., Di Francesco, S. A., Mastroberardino, S., & Macaluso, E. (2015). Parietal cortex integrates contextual and saliency signals during the encoding of natural scenes in working memory. *Human Brain Mapping*, 36, 5003-5017.
- Santangelo, V., & Macaluso, E. (2013). Visual salience improves spatial working memory via enhanced parieto-temporal functional connectivity. *Journal of Neuroscience*, 33, 4110-4117.

- Santangelo, V., Olivetti Belardinelli, M., & Spence, C. (2007). The suppression of reflexive visual and auditory orienting when attention is otherwise engaged. *Journal of Experimental Psychology: Human Perception and Performance*, 33, 137-148.
- Santangelo, V., & Spence, C. (2008). Is the exogenous orienting of spatial attention truly automatic? Evidence from unimodal and multisensory studies. *Consciousness and Cognition*, 17, 989-1015.
- Scerif, G. (2010). Attention trajectories, mechanisms and outcomes: at the interface between developing cognition and environment. *Developmental Science*, 13, 805-812.
- Schwartz, S., Vuilleumier, P., Hutton, C., Maravita, A., Dolan, R. J., & Driver, J. (2005). Attentional load and sensory competition in human vision: modulation of fMRI responses by load at fixation during task-irrelevant stimulation in the peripheral visual field. *Cerebral Cortex*, 15, 770-786.
- Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics*, 59, 1-22.
- Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica*, 135, 77-99.
- Treisman, A. M., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12, 97-136.
- Trick, L. M., & Enns, J. T. (1998). Lifespan changes in attention: The visual search task. *Cognitive Development*, 13, 369-386.
- Underwood, G., Foulsham, T., van Loon, E., Humphreys, L., & Bloyce, J. (2006). Eye movements during scene inspection: A test of the saliency map hypothesis. *European Journal of Cognitive Psychology*, 18, 321-342.
- Van der Burg, E., Olivers, C. N., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 1053-1065.

Vu, K.-P. L., Strybel, T. Z., & Proctor, R. W. (2006). Effects of displacement magnitude and direction of auditory cues on auditory spatial facilitation of visual search. *Human Factors, 48*, 587-599.

Waszak, F., Li, S.-C., & Hommel, B. (2010). The development of attentional networks: Cross-sectional findings from a life span sample. *Developmental Psychology, 46*, 337-349.

Footnotes

1. The main reason for using the auditory modality for spatial cuing was to investigate simultaneously both auditory and visual distraction, implemented respectively in terms of “invalid” audio-spatial cues and visual distractors (i.e., non-target objects) in the scene. Future studies might verify whether visuospatial cues – presented on the periphery of the scene or cuing within-scene locations – have a different impact on searching performance on complex visual scenes.
2. For completeness we also report post-hoc analyses exploring the Age group x Cue type interaction on raw RT data. Here we conducted a one-way repeated measure ANOVA with the within-participants factor of Cue type (valid, neutral or invalid) for each group. All three ANOVAs were significant: 6 y.o. [$F(2, 78) = 14.5, p < .001, \eta^2 = .270$]; 10 y.o. [$F(2, 74) = 15.9, p < .001, \eta^2 = .300$]; 20 y.o. [$F(2, 60) = 6.6, p = .003, \eta^2 = .181$]. Post-hoc comparisons indicated faster searching performance following valid than neutral auditory cues (i.e., orienting effects) in all the three age groups: 6 y.o., magnitude of 21 ms, $p = .044$; 10 y.o., 16 ms, $p = .010$; 20 y.o., 13 ms, $p = .015$. In contrast, searching performance was slower following invalid than neutral auditory cues (i.e., reorienting effect) in the 6 (33 ms, $p = .002$) and 10 (18 ms, $p = .004$) y.o. groups, but not in the 20 y.o. group (5 ms, $p = .305$).
3. Again for completeness we report post-hoc analyses exploring the Age group x Cue type interaction on accuracy scores. Again three one-way ANOVAs (Cue type: valid, neutral, invalid), were conducted, one for each Age group. The ANOVAs of the 6 and 10 y.o. groups were significant ([$F(2, 78) = 7.8, p = .001, \eta^2 = .168$] and [$F(2, 74) = 7.5, p = .001, \eta^2 = .169$], respectively), indicating reorienting effects in both groups, that is a specific decrease of accuracy following invalid than neutral cues (6 y.o.: magnitude of 2.2%, $p = .002$; 10 y.o.: 1.9%, $p = .006$), that not differed instead from valid cues (6 y.o.: 0.3%, $p = .650$; 10 y.o.: 0.5%, $p = .403$). By

contrast, the ANOVA of the 20 y.o. group was not significant, [$F(2, 60) < 1$, n.s., $\eta^2 = .013$], indicating similar accuracy irrespective of auditory cue types.

Table 1. Participant characteristics for each age group, and magnitude of orienting (“valid minus neutral” trials) and reorienting (“invalid minus neutral” trials) in terms of both reaction time (RT) and accuracy (ACC) data \pm standard error of the mean in the three age groups.

	6 y.o. n = 40 (15 f, 25 m) Mean age (years) = 6.1 (SD = 0.3, Range 5-7)		10 y.o. n = 38 (21 f, 17 m) Mean age (years) = 10.3 (SD = 0.4, Range 10-11)		20 y.o. n = 31 (14 f, 17 m) Mean age (years) = 20.0 (SD = 1.6, Range 18-23)	
	Orienting	Reorienting	Orienting	Reorienting	Orienting	Reorienting
RT (ms)	-21 \pm 10	33 \pm 9	-16 \pm 4	18 \pm 6	-13 \pm 4	5 \pm 6
ACC (%)	0.3 \pm 0.5	-2.2 \pm 0.8	0.5 \pm 0.4	-1.9 \pm 0.8	0.4 \pm 0.4	0.3 \pm 0.5

Table 2. Standardized regression coefficients (betas), t-values, significance (p) and effect size (η^2) resulted from the multivariate multiple regression (MMR) models showing the impact of predictors (TgS and NSL) on the three dependent variables (searching following valid, neutral or invalid cues) in the three age groups.

Group	Dependent	Predictor	Beta	t-value	p	η^2
6 y.o.	Val	TgS	-0.090	-.602	.549	.005
		NSL	8.628	2.715	.008	.099
	Neu	TgS	-0.063	-.412	.681	.003
		NSL	5.201	1.601	.114	.037
	Inv	TgS	-0.185	-1.021	.311	.015
		NSL	9.967	2.597	.012	.091
10 y.o.	Val	TgS	-0.032	-.402	.689	.002
		NSL	5.776	3.470	.001	.152
	Neu	TgS	-0.014	-.156	.876	.000
		NSL	2.457	1.288	.202	.024
	Inv	TgS	-0.050	-.635	.528	.006
		NSL	4.262	2.543	.013	.088
20 y.o.	Val	TgS	-0.167	-2.360	.021	.077
		NSL	2.097	1.403	.165	.029
	Neu	TgS	-0.056	-.729	.469	.008
		NSL	3.042	1.865	.067	.049
	Inv	TgS	-0.109	-1.543	.127	.034
		NSL	2.608	1.739	.087	.043

Note: Significant effects are marked with bold.

Figure captions

Figure 1. A) Diagram showing the sequence of events during one block of trials. This began with the presentation of a target display defining the to-be-searched target in the following scene. After any response button press, a sequence of nine scenes was then presented. Each scene was displayed for 3000 ms, with an inter-stimulus interval of 2000 ms. Each picture was preceded by a 50 ms burst of white noise presented on either the left, right, or both hemifields. Participants' pressed one of two response buttons according to left vs. right location of the to-be-searched target. B).Saliency map of an example scene. The map highlight several salient areas, each characterized by a saliency peak. Values of saliency peaks change according to the color map on the right. C) Mean reaction time (RT; left graph) and mean accuracy (right graph) of the different age groups during visual search following valid, neutral or invalid auditory cues. The error bars represent the standard error of the mean. D) Standardized regression coefficients (beta values) derived from the multivariate multiple regression (MMR) models used to assess the predictivity of number of salient locations (NSL) and target saliency (TgS) on search performance following valid, neutral or invalid auditory cues in the three age groups (see also Tab. 2). The error bars represent the standard error of the mean.

Figure 1

