

Expert Attention: Attentional allocation depends on the differential development of multisensory
number representations

Pawel J. Matusz^{1,2}, Rebecca Merkley³, Michelle Faure⁴ and Gaia Scerif⁴

¹ Information Systems Institute, University of Applied Sciences Western Switzerland
(HES-SO Valais), Sierre, Switzerland

² The LINE (Laboratory for Investigative Neurophysiology), Neuropsychology and Neurorehabilitation
Service and Radiodiagnostic Service, University Hospital Center and University of Lausanne,
Lausanne, Switzerland

³ Numerical Cognition Laboratory, University of Western Ontario, Canada

⁴ Attention, Brain and Cognitive Development Group, University of Oxford, United Kingdom

Corresponding authors:

Pawel Matusz pawel.matusz@gmail.com or Gaia Scerif gaia.scerif@psy.ox.ac.uk

Traditional models developed within cognitive psychology suggest that attention is deployed flexibly and irrespective of differences in expertise with to-be-attended stimuli. However, everyday environments are inherently multisensory and observers differ in familiarity with particular unisensory representations (e.g., number words, in contrast with digits). To test whether the predictions of the traditional models extend to such naturalistic settings, six-year-olds, 11-year-olds and young adults (N=83) searched for predefined numerals amongst a small or large number of distractor digits, while distractor number words, digits or their combination were presented peripherally. Concurrently presented number words and audiovisual stimuli that were compatible with the target digit facilitated young children's selective attention. In contrast, for older children and young adults number words and audiovisual stimuli that were incompatible with their visual targets resulted in a cost on reaction time. These findings suggest that multisensory and familiarity-based influences interact dynamically as they shape selective attention. Therefore, models of selective attention should include multisensory and familiarity-dependent constraints: **more or less familiar object representations across modalities will be attended to differently, with their effects visible as predominant benefits for attention at one level but costs at another.**

Expert Attention: Attentional allocation depends on the differential development of multisensory number representations

Goal-directed behaviour requires allocating attentional resources preferentially to the stimuli currently important to the task at hand (*selective attention*). The dominating models portray selective attention as a flexible resource, deployed similarly across distinct types of stimuli and contexts (Bundesen, Habekost, & Kyllingsbaek, 2005; Duncan, 2010; Lavie, 2010). Despite increasing support for the importance of stimulus sensory modality (Iordanescu et al., 2010; Matusz & Eimer, 2011, 2013; Sarmiento et al. 2016) and one's expertise (Wu et al., 2015) as sources of top-down attentional control, the majority of these models have been tested using exclusively unimodal (typically visual) and highly familiar stimuli. Here, we utilised naturally occurring age-related differences in expertise for specific naturalistic stimuli (i.e., auditorily-presented number words versus visually presented digits) to investigate the influence of differences in familiarity with to-be-attended stimuli on the deployment of selective attention in naturalistic, multisensory environments.

Some fundamental mechanisms controlling attention have been clarified by studying the effects of "perceptual load" (Lavie, 1995, 2010). In a typical paradigm (Lavie & Cox, 1997), participants search for one of two target stimuli (X or N) within a search array always accompanied by a peripheral distractor. If the stimulus array is small, the peripheral distractor interferes with search, for example, slowing it down and speeding it up depending on whether distractors and targets match in identity (*stimulus-response compatibility effects*). In larger arrays, compatibility effects become attenuated and/or eliminated. The increasing number of to-be-processed stimuli in the array is argued to deplete the available attentional resources. These processes show a protracted developmental trajectory (Couperus, Hunt, Nelson, & Thomas, 2011; Huang-Pollock, Carr, & Nigg, 2002; Lavie, 2010), and, critically, are modulated by whether targets and distractors engage different senses. For example, visual task demands were initially reported to attenuate cross-modal distraction, and this is in turn interpreted in terms of attentional resources being shared across the

senses (Macdonald & Lavie, 2011; Raveh & Lavie, 2015). Conversely, a growing body of research has demonstrated differing effects for distractors in distinct sensory modalities (e.g., Mahr & Wentura, 2014; Sugimoto & Katayama, 2017; Tellinghuisen & Nowak, 2003), which suggest that attentional resources are indeed allocated differently across senses. Surprisingly, perceptual load effects have rarely been tested against multisensory stimuli, despite the ubiquity of the latter in naturalistic environments. Yet, here, results are more consistent. Distraction elicited by multisensory stimuli in visual search tasks is not reduced by increasing attentional demands (in contrast with decreased distraction with increases in visual “perceptual load”, Matusz et al., 2015; see also Santangelo & Spence, 2007).

More broadly, multisensory information is well-known to benefit neural processing and behaviour, as the brain integrates information from different senses (Murray & Wallace, 2012). These benefits are increasingly understood to rely on distinct multisensory processes, varying in their dependence on top-down attentional control but also on the current contents of long-term memory. Evidence from infant studies suggests that multisensory processes, even if rudimentary, are present soon after birth and dominate early attentional control (Bahrick, Lickliter, Castellanos, & Todd, 2015; Lewkowicz, Leo, & Simion, 2010; Neil, Chee-Ruiter, Scheier, Lewkowicz, & Shimojo, 2006). For objects with which participants become highly familiar early in life, developmentally informed designs have offered a unique window into understanding how the developing ability to deploy attention enables one to ignore distraction. For example, Matusz et al. (2015) adapted Lavie and Cox’s (1997) task to demonstrate that young children – but not older children or adults – are paradoxically “shielded” from distraction by colour label-shape pairings when their primary attention task is difficult.

In everyday situations, stimuli are not only typically complex, but also differ in the level of expertise that observers have with them. What is the influence of expertise on attention allocation in naturalistic, multisensory settings? Differential experience with stimuli in our environment might be one top-down factor that controls the influence of both stimulus type and of an individual’s age

on attentional control. Indeed, previous unisensory research points to the importance of object expertise in controlling attention deployment (Wu et al. 2015). Specifically, having more experience with a given stimulus or category may allow for a more efficient attentional search template (Wu et al., 2015). Similarly, a long tradition of contextual cueing and memory-guided attention paradigms demonstrate that our past experiences can shape the way in which we filter distractions to deploy selective attention (e.g., Chun & Jiang, 1998 Goldfarb, Chun, & Phelps, 2016; Summerfield, Lepsien, Gitelman, Mesulam, & Nobre, 2006). As a whole, this work suggests that differential expertise with to-be-attended stimuli, be they the targets of attention or distractors, can influence how attention is deployed.

The influence of stimulus familiarity might be even more crucial in multisensory environments, where expertise with one stimulus format frequently develops before that with another, as is the case with number words and numerals. Studying attention allocation towards digits can offer a unique window into the role of such typical asymmetries in stimulus familiarity as a source of top-down attentional control. Digit labels (number words) are learnt earlier than written numerals, which are learnt more systematically and to a high degree of fluency only after formal schooling begins (e.g., Bloom & Wynn, 1997; Davidson, Eng, & Barner, 2012; Lyons, Price, Vaessen, Blomert, & Ansari, 2014; Merkley & Ansari, 2016; Wagner & Johnson, 2011). Specifically, children learn that number words refer to specific quantities before learning that Arabic digits also refer to specific quantities, and they integrate the multiple representations of numbers over the first years of school (Hurst, Anderson, & Cordes, 2017; Jiménez Lira, Carver, Douglas, & LeFevre, 2017). Furthermore, while older children and adults' ability to judge the size of Arabic digits is disrupted by digits' associated numerical magnitude, young children do not demonstrate these interference effects (Rubinsten, Henik, Berger, Shahar-Shalev, 2002). Therefore, just as young children cannot experience conflict on the classic Stroop task until they are proficient readers, children also do not seem to experience conflict on a numerical size Stroop task until they reach a certain level of proficiency with number symbols. This highlights how attentional allocation may be influenced by

familiarity with stimuli, in this case with number symbols. Moreover, multisensory processes specific to letters or numbers develop only following substantial experience with each unisensory format (Froyen, Bonte, van Atteveldt, & Blomert, 2009). Until such expertise is obtained, how attention is allocated to different representational formats is likely to differ both in magnitude and in valence (i.e., in whether it accrues costs or benefits). Dominating models portray selective attention as relatively unconstrained by the multisensory nature of the to-be-attended material.

Here, we tested children that recently entered primary school, as well as older children and adults on visual search for digits, and we measured traditional attentional benefits and costs of visual, auditory and audiovisual distractors. We predicted that the specific effects of number words and digits on selective attention and their benefits versus costs on visual search, would differ across the three age groups that differ in stimulus familiarity with numerical representation formats.

METHODS

Participants

Participants were 28 school children in at the end of **Year 1** (5-6 year-olds), 29 children in Year 6 (10-11 year-olds) and 30 adults. Four participants were excluded on the basis of their overall accuracy being lower than 55% correct (3 children in **Year 1**, 1 child in Year 6). The final sample consisted of 25 “6-year-olds” (aged 5.1–6.3 years, mean \pm SD = 5.7 \pm 0.4 years; 15 females), 28 “11-year-olds” (aged 10.1–11.1 years, mean \pm SD = 10.6 \pm 0.3; 12 females), and 30 young-adult participants (aged 19.3–24.8 years; mean \pm SD = 22 \pm 1.4, 18 females). This sample size was chosen with the aim to be a comparable sample size to a previous similar study from our group (Matusz et al., 2015). A post hoc power analysis revealed that the power for testing our hypothesized three-way interaction in the current study was .99. Adults were undergraduate students, recruited via email. Children in **Year 1** and Year 6 were recruited **in the late summer term** from nearby primary schools, and **they all took part in the study over the first three months of the school year**. All adult participants provided informed consent, and parental consent was obtained for each child before starting the experiment,

according to the procedures set out by the appropriate Ethics Review Board. Children were rewarded with items of stationery (stickers, pencils, etc.), whereas adults participated without compensation. All participants had normal hearing, and normal or corrected-to-normal vision.

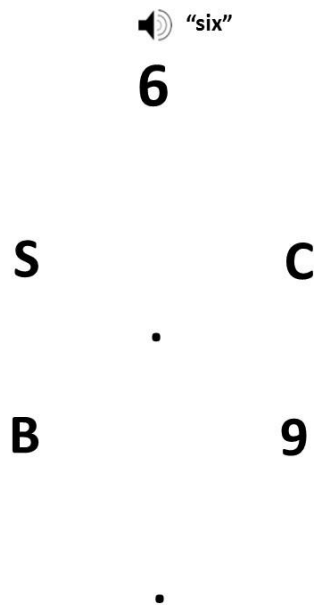


Figure 1. An example of a search display in which the target stimulus ‘9’ was presented at high load, with a target-incompatible, audiovisual distractor.

Stimuli and Procedure

Stimuli were created and presented using E-Prime v.2 (Psychology Software Tools, Pittsburgh, PA) on a Dell Latitude D430 laptop positioned approximately 50 cm from the participants. Each trial began with 1000-ms-long central fixation point, immediately followed by a 200-ms-long visual search display, followed by a 5000ms blank white response screen. The task was set in the context of helping Santa to sort presents in time for Christmas Eve, to make the task more engaging to the young participants. As shown in [Figure 1](#), participants searched for one of two possible digits (“6” or “9”) and pressed one of two mouse buttons upon detection (left and right, respectively). In set-size 4

blocks, the target ($0.6^\circ \times 0.6^\circ$) was surrounded by three letters, “B”, “C”, and “S” ($0.6^\circ \times 0.6^\circ$), appearing randomly and equiprobably in one of four possible locations along the circumference of a circle centered at fixation (2.1° radius). In set-size 1 blocks, the target was presented alone at one of the six locations. The nontarget letters were chosen for their curved appearances, making them perceptually similar to the target digits. **Set-size fixed within each block because prior work investigating the differential effects of perceptual load in children compared to adults (Huang-Pollock et al., 2002) in a version of the perceptual paradigm that randomize set-size across trials identified very poor performance in young children, with 25% of children not meeting a 67% overall accuracy criterion. This was likely due to confusion. Here, we opted for blocking set-size, as it was a previously studied variable, while randomizing modality and the level of familiarity associated with modality, our two novel target variables.**

The visual peripheral distractor was a larger-size “6”, “9” or “3” ($0.8^\circ \times 0.8^\circ$) presented at a distance of 4.1° from the fixation point, presented on the central midline, rather than to the right or left of fixation, to avoid inducing incompatibility effects of specific digits with number line position. The “3” represented a condition neutral in terms of congruence with the target digit, and, like search-array nontarget letters, was chosen for its curved appearance. Auditory distractors were voice recordings (500ms) of the words “six”, “nine”, “three”, each lasting 500ms, produced using GarageBand 5.1. In the audiovisual distraction condition, the distractors were presented visually along the vertical meridian, above or below the fixation, and concurrently centrally through the speaker. On a given trial, either distractor type either matched the identity of the current target (identity of the target was randomly chosen on each trial), matched identity of the other target or matched identity of neither target - target–distractor “compatible”, “incompatible” and “neutral” trials, respectively.

The experiment lasted 30 minutes, deemed a length appropriate for the youngest children following a pilot study. Adults were tested in a quiet room at the university’s Psychology department, and children in a quiet room located in their school. An experimenter sat next to

participants throughout the whole experiment. Participants completed first a slow-practice block familiarising them with the task and the different trial types (9 trials: 5 set-size 1, 4 set-size 4). These trials consisted of a fixation point (1000 ms), followed by search display (infinite duration), then a feedback screen involving a happy/sad emoji (500 ms). Participants then completed 2 fast-practice blocks (each of 9 trials), following the experimental trial structure described above. Subsequently, eight experiment blocks of 36 trials were presented, four blocks for each of two set-size conditions. Within each block, there were 12 trials with visual distractors, 12 with auditory distractors, and 12 with audiovisual distractors (2 repetitions for each of the two target identities and three compatibility trials). This resulted in a total of 288 trials. Each block started with a written task instruction, with no time limit. Participants were instructed to respond as quickly and accurately as possible. Feedback was provided immediately after the response, in the form of a happy/sad face emoji (500-ms duration). If no response was made, a “Faster!” message appeared (also 500-ms duration). Each block finished with a screen presenting to the participant his block and cumulative accuracy in encouraging fashion. The experiment was concluded with an overall accuracy score and thanks to the participant for “saving Christmas”.

Data analysis

To identify the attentional effects triggered by different distractor formats as a function of stimulus familiarity, we analysed medians of correct reaction time (RT) in a four-way mixed analysis of variance (ANOVA), with compatibility (distractor compatible vs. neutral vs. incompatible with the target identity), set-size (1 versus 4 items in the search array), and distractor type (visual versus auditory versus audiovisual) as within-subjects factors, and age (adults vs. 11-year-olds vs. 6-year-olds) as a between-subjects factor. “Attentional benefits” (i.e., **faster RTs on compatible vs. neutral trials**) and “attentional costs” (i.e., **slower RTs on incompatible vs. neutral trials**) and how they were modulated by distractor type, set-size and age of observer were of primary interest. Thus, the follow-up analyses focused on these attentional effects, where appropriate. As error rate data

violated the assumptions of parametric statistics, we report only simple effects (attentional costs and benefits), without full statistical analyses.

RESULTS

Age influences attentional effects in multisensory settings

Our analyses revealed a number of higher-order interactions (described in full in the Supplemental Online Materials, “SOM” henceforth). Most importantly, the pattern of attentional effects was determined simultaneously by the type of distractor and observer’s age, as demonstrated by a three-way Compatibility x Distractor type x Age interaction, $F(8,320) = 3.16$, $p = .002$, $\eta_p^2 = .073$. This interaction was not further modulated by Load, $F < 1$. For distractor number words, 6-year-olds demonstrated exclusively attentional benefits (93ms, $p = .015$; costs: $p > .55$), while 11-year-olds and adults - exclusively attentional costs (28ms, $p = .01$, and 15ms, $p = .007$, respectively; benefits: p ’s $> .17$; Figure 2). No age group showed reliable visually-induced attentional effects: p ’s $> .53$ in 6-year-olds, p ’s $> .19$ in 11-year-olds, and p ’s $> .34$ in adults. For audiovisual distractors, again 6-year-olds showed attentional benefits (136ms, $p = .009$; costs: $p > .12$), while adults demonstrated both attentional benefits (16ms, $p = .012$) and costs (23ms, $p = .002$); in 11-year-olds, neither costs, nor benefits were reliable, p ’s $> .13$.

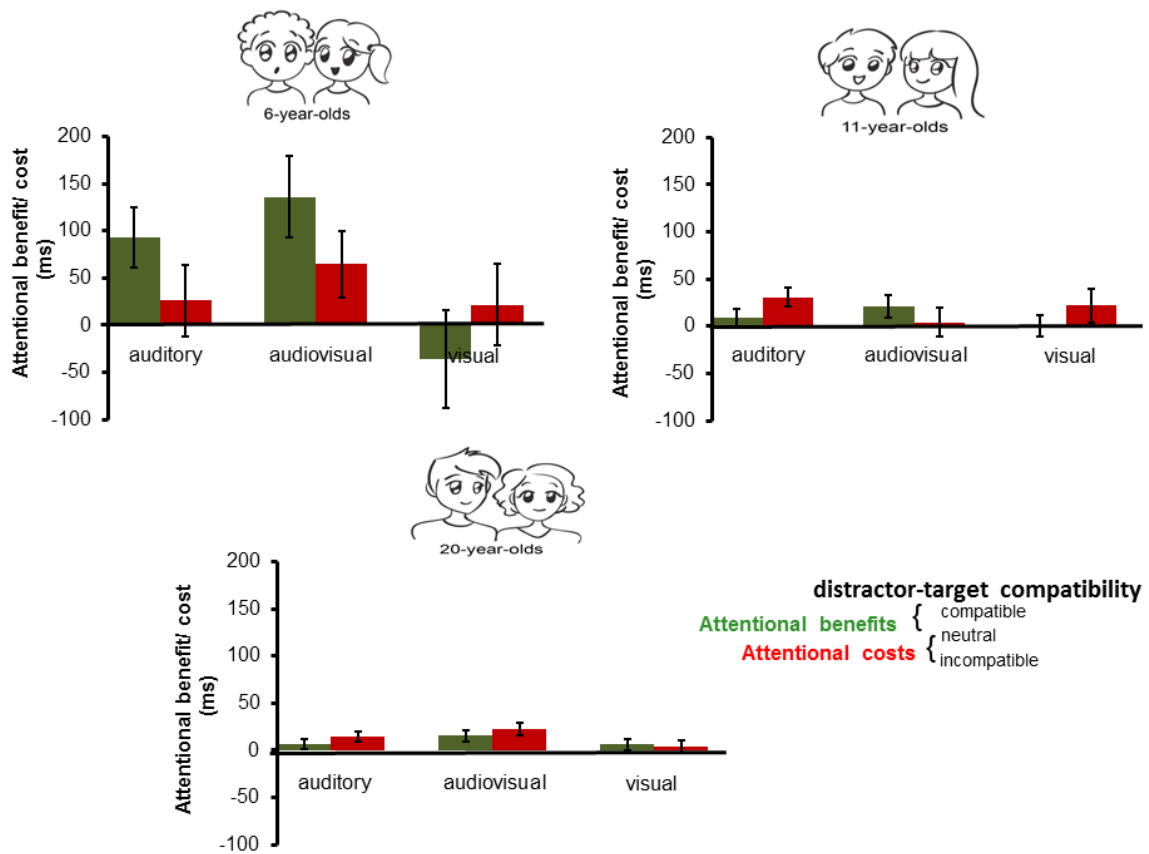


Figure 2. Attentional costs (in red) and attentional benefits (in green) elicited by the three distractor types in 20-year-olds, 11-year-olds and 6-year-olds.

Age influences the load-induced effects of multi-sensory distractors on visual search

The extent to which search speed was impacted by set-size and the type of the distractor accompanying the search array was also determined by age, as evidenced by a three-way Age x Distractor type x Set size interaction, $F(4,160) = 3.32, p = 0.012, \eta_p^2 = .077$. Specifically, follow-up analyses using two-way Distractor type x Set-size ANOVAs in each age group revealed that only 6-year-olds showed evidence of preferential processing of number words. **These results were supported by a main effect of set-size, $F(1,24) = 29.1, p < .001, \eta_p^2 = .55$, and an interaction with distractor type, $F(2,48) = 3.69, p = .032, \eta_p^2 = .13$.** That is, at set-size 1, search for 6-year-olds was faster when accompanied by number words (1208ms) than numerals (1275ms), $p < .001$, in turn faster than trials with audiovisual distractors (1340ms), $p < .03$ (see Figure 3). At set-size 4, search was similar across the 3 distractor type trials, p 's $> .37$.

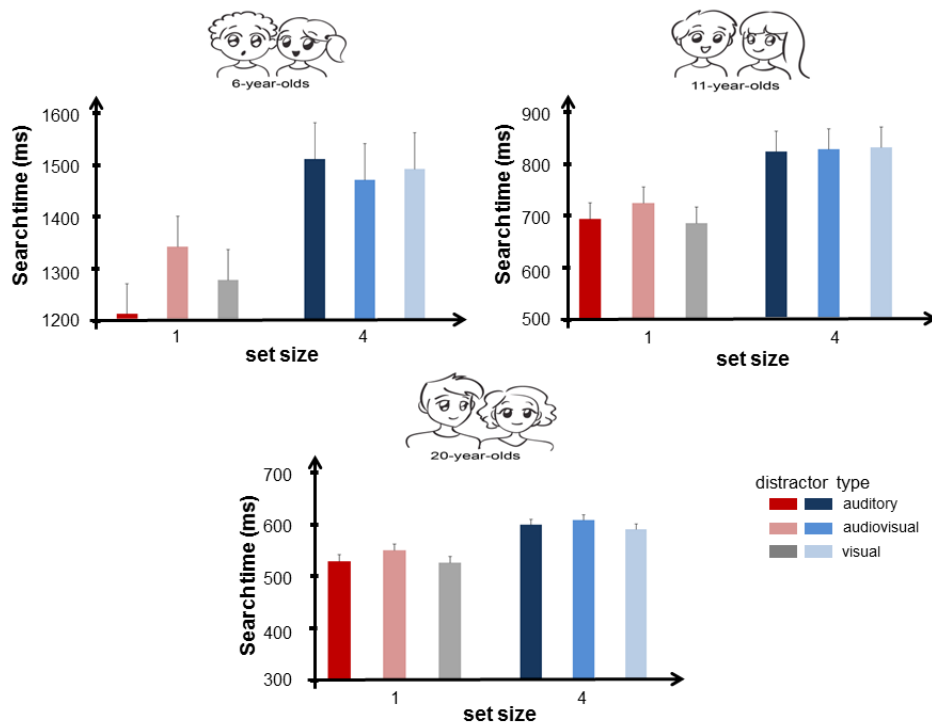


Figure 3. Mean search RTs across smaller and larger set-size when search was accompanied by an auditory, visual and audiovisual distractor, shown separately for 20-year-olds, 11-year-olds and 6-year-olds.

In contrast, in 11-year-olds, a main effect of set-size, $F(1,27) = 60.32, p < .001, \eta_p^2 = .69$, was modulated by distractor type, $F(2,54) = 4.13, p = .02, \eta_p^2 = .13$, but this time search speed at set-size 1 was similar across numeral (693ms) and number word distractors (701ms), $p > .41$, with slowest responses again on audiovisual trials (732ms), $p's = .002$. As before, at set-size 4 search times was not modulated by distractor type, $p's > .73$. In adults, as in the other two groups, search was slower with increased set-size (534ms vs. 607ms), $F(2,58) = 60.32, p < .001, \eta_p^2 = .69$. Overall, adult search was also similar on trials with numerals (561ms) and number word distractors (567ms), $p = .2$, and audiovisual distractors triggering the slowest responses (584ms), $p's > .002$. In adults, the two factors did not interact $F(2,58) = 1.46, p = .25, \eta_p^2 = .05$.

Accuracy data

Error rates data failed to fulfil the parametric test assumptions. Wilcoxon's signed rank tests showed in 6-year-olds: costs for number words at set-size 4, 13%, $W=35$, $p=.003$, costs for numerals at set-size 1, 7%, $W=62$, $p=.037$, and costs for **numeral-word audiovisual pairings** at set-size 1, 8%, $W=52.5$, $p=.016$; for audiovisual pairings, there was also a non-statistical trend for benefits at set-size 4, 6%, $W=74.5$, $p=.055$. No reliable error-rate costs or benefits were found in adults (all p 's $>.4$) or 11-year-olds (all p 's $>.2$, with the exception of borderline statistically significant benefits for number words at set-size 1, 6%, $W=94$, $p=.051$).

DISCUSSION

Our study aimed to test whether the predictions of predominant attentional theories like perceptual load theory (e.g., Lavie, 1995, 2005) extend to contexts of multisensory stimuli and stimuli whose unisensory representations develop at different time-frames. **These aims are important, because in everyday environments selective attention is challenged by multisensory input, so testing unisensory stimuli only is not sufficient. In addition, our stimulus familiarity as observers varies for distinct categories and representational formats of to-be-attended stimuli.** Indeed, studies of contextual cueing and of memory-guided attention have demonstrated how prior experience influences the deployment of selective attention (e.g. Chun & Jiang, 1998; Goldfarb et al., 2016). However, this role of differential familiarity with stimuli has not to our knowledge been extended to test age-related differences in selective attention for stimuli in distinct sensory modalities.

A number of novel findings emerged. First, attentional costs and benefits differed depending on the distractor as well as the age of the observer. Indeed, numbers presented in the auditory modality (number words), the sensory modality most familiar to young observers, facilitated their visual selective attention, whereas they hindered visual selective attention in older observers. These findings are especially interesting because symbolic representations of number vary in the extent to which they are well established by the time of school entry, as in the case of the contrast between number words and written digits (Jiménez Lira et al., 2017), or the contrast between physical size

and numerical magnitude (Rubinsten et al., 2002). Number word representations are acquired and established over the preschool years (Bloom & Wynn, 1997; Davidson et al., 2012), whereas fluency in identifying and operating on visually presented Arabic numerals is slower to emerge and does so during the first years of formal school (Lyons et al., 2014; Reynvoet, De Smedt, & Van den Bussche, 2009). It is therefore very interesting that the effects of number words as distractors are apparent and more robust than those of visual distractor digits, and also that they move from benefits to costs from early childhood into adulthood. These changing benefits and costs may be accounted for by changes in proficiency with digits: as initially digit representations are not as well established, number words help identification in early childhood, whereas in older children and adults number words simply bear a cost. **It is worth providing more details on the maths education context surrounding our participating children. Although the youngest children in our study were young, they had just completed what in the United Kingdom is defined as the Early Years Foundation Stage, which sets as a target the ability to count and recognise numbers up to 20 (Standards & Testing Agency, 2013). This entails that they are at a stage of development when they have been exposed to basic numerals for a significant period, which makes our findings more striking: even though they count up to twenty really readily, when young children are required to select numerals under speeded conditions, the modality in which additional information is provided matters a great deal, with number words, their most familiar numerical representation, accruing a real benefit. In contrast, our older children would have typically transitioned to retrieving and operating on multi-digit numerals, working under speeded conditions, a transition away relying on number words alone.** Of note, it may be tempting to dismiss these findings as dependent on simply poorer attentional capacity by the youngest children: however, the finding that younger children do not simply experience greater interference from all distractors, but benefit selectively from being presented with additional number words, suggests a much more nuanced interplay between prior experience, multi-sensory differences and attentional selection.

In turn, these findings highlight the need for attentional models to extend to multi-sensory stimuli with which we are differentially familiar. In addition to these theoretical implications, the current findings highlight a gap between cognitive models of attention that attempt to explain attention to everyday environments: for example, existing cognitive models may bear little relevance to the classroom, because of the narrow focus on certain categories of visual stimuli and tightly controlled experimental environments. These are ultimately very different from complex, multi-sensory classroom environments. Simultaneously, emulating the demands of these environments in the lab might become an important tool in understanding neurocognitive functions, such as selective attention and its development, as they occur in the real world (Matusz, Dikker, Huth, & Perrodin, 2018). Secondly, auditory distractor number words affected all participants, irrespective of how demanding the central load of the task was, consistent with a growing body of work using both simple and complex stimuli (Mahr & Wentura, 2014; Sugimoto & Katayama, 2017; Tellinghuisen, Cohen, & Cooper, 2016; Tellinghuisen & Nowak, 2003; Zäske, Perlich, & Schweinberger, 2016). More broadly, these findings therefore support a view of attentional resources that are differentially engaged by stimuli presented in different modalities, rather than a single and limited central resource (Matusz et al., 2015; Tellinghuisen & Nowak, 2003; Welch & Warren, 1980).

Finally, multisensory representations of those numbers (i.e., concurrently presented digits and number words) resulted in effects on visual selective attention that exceeded those of unisensory stimuli. For the youngest children, these audiovisual stimuli resulted in benefits, parallel to those caused by number words alone. For young adults, both benefits and costs were apparent. These findings are consistent with studies investigating multisensory distraction with simpler stimuli of similar unisensory levels of stimulus familiarity (e.g., simple shapes and colours, Matusz et al., 2015): multisensory distractors have a robust effect on visual attentional selection, but an effect that is modulated by the age of observers.

Amongst the limitations, it is worth noting that it is difficult to explain unclear effects of perceptual load here (cf. Huang-Pollock et al., 2002; Matusz et al., 2015). However, these may depend on the dominance and developmental primacy of the auditory distractor modality measured here, compared to the stimuli presented in those studies. Furthermore, here we only employed number stimuli. It would be very interesting to discuss possible parallels between audiovisual number versus letter-speech sound representations. Indeed, Huang-Pollock et al. (2002) employed visual letter stimuli in their adaptation of the perceptual load paradigm. Performance by the youngest children in their sample was very poor, but it would be of great interest to test how this performance might be improved / modified by presenting speech sound representations along-side those letter stimuli. The hypothesis emerging from our data would be that the youngest children would benefit from concurrent speech sound presentation, if they were required to select and identify one of two target letters in an array of distractors. Particularly interesting would be highly confusable letters (such as “b” and “d”, or “p” and “q”), for which inversion mistakes in early readers are very common. We have now included a discussion of this point and, as we are not reading experts, most definitely hope that the suggestion for future directions will be tackled by readers.

To conclude, attentional selection theories have not previously focused on the extent to which prior stimulus familiarity influences the interaction between multisensory processes and selective attention (e.g. modulating attention under high task demands). Here we highlight the relevance of both multisensory processes and stimulus familiarity to understanding selective attention in the natural world where these variabilities are typically at play.

Authorship information

All authors contributed to the study design. M.F. led on data collection. P.J.M. led in on data analysis with input from G.S. and R.M. on writing and modifications of manuscript drafts. P.J.M., R.M. and G.S. all contributed to the final manuscript.

Competing Interests

The authors declare no competing financial interests.

Acknowledgements

PJM is funded by the Pierre Mercier Foundation, Fondation Asile des Aveugles and the Swiss National Science Foundation (grant PZ00P1_174150). G.S. was supported by a 2012 Scholar Award (Understanding Human Cognition) of the James S. McDonnell Foundation.

REFERENCES

- Bahrick, L. E., Lickliter, R., Castellanos, I., & Todd, J. T. (2015). Intrасensory Redundancy Facilitates Infant Detection of Tempo: Extending Predictions of the Intersensory Redundancy Hypothesis. *Infancy*, 20(4), 377–404. <https://doi.org/10.1111/infa.12081>
- Bloom, P., & Wynn, K. (1997). Linguistic cues in the acquisition of number words. *Journal of Child Language*, 24(3), 511–533. <https://doi.org/10.1017/S0305000997003188>
- Bundesen, C., Habekost, T., & Kyllingsbaek, S. (2005). A neural theory of visual attention: bridging cognition and neurophysiology. *Psychological Review*, 112(2), 291–328. <https://doi.org/10.1037/0033-295X.112.2.291>
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, 36(1), 28–71. <https://doi.org/10.1006/cogp.1998.0681>
- Couperus, J. W., Hunt, R. H., Nelson, C. A., & Thomas, K. M. (2011). Visual search and contextual cueing: Differential effects in 10-year-old children and adults. *Attention, Perception, and Psychophysics*, 73(2), 334–348. <https://doi.org/10.3758/s13414-010-0021-6>
- Davidson, K., Eng, K., & Barner, D. (2012). Does learning to count involve a semantic induction? *Cognition*, 123(1), 162–173. <https://doi.org/10.1016/j.cognition.2011.12.013>

- Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, 14(4), 172–179.
<https://doi.org/10.1016/j.tics.2010.01.004>
- Froyen, D. J. W., Bonte, M. L., van Atteveldt, N., & Blomert, L. (2009). The Long Road to Automation: Neurocognitive Development of Letter–Speech Sound Processing. *Journal of Cognitive Neuroscience*, 21(3), 567–580. <https://doi.org/10.1162/jocn.2009.21061>
- Goldfarb, E. V., Chun, M. M., & Phelps, E. A. (2016). Memory-Guided Attention: Independent Contributions of the Hippocampus and Striatum. *Neuron*, 89(2), 317–324.
<https://doi.org/10.1016/j.neuron.2015.12.014>
- Huang-Pollock, C. L., Carr, T. H., & Nigg, J. T. (2002). Development of selective attention: Perceptual load influences early versus late attentional selection in children and adults. *Developmental Psychology*, 38, 363–375. <https://doi.org/10.1037//0012-1649.38.3.363>
- Hurst, M., Anderson, U., & Cordes, S. (2017). Mapping Among Number Words, Numerals, and Nonsymbolic Quantities in Preschoolers. *Journal of Cognition and Development*, 18(1), 41–62.
<https://doi.org/10.1080/15248372.2016.1228653>
- Iordanescu, L., Grabowecky, M., Franconeri, S. L., Theeuwes, J., & Suzuki, S. (2010). Characteristic sounds make you look at target objects more quickly. *Attention, Perception & Psychophysics*, 72(7), 1736–1741. <https://doi.org/10.3758/APP>
- Jiménez Lira, C., Carver, M., Douglas, H., & LeFevre, J. A. (2017). The integration of symbolic and non-symbolic representations of exact quantity in preschool children. *Cognition*, 166, 382–397.
<https://doi.org/10.1016/j.cognition.2017.05.033>
- Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 451–468.
<https://doi.org/10.1037/0096-1523.21.3.451>
- Lavie, N. (2005). Distracted and confused? Selective attention under load. *Trends in Cognitive Sciences*, 9, 75–82. <https://doi.org/http://dx.doi.org/10.1016/j.tics.2004.12.004>

- Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current Directions in Psychological Science*, 19, 143–148. <https://doi.org/10.1177/0963721410370295>
- Lavie, N., & Cox, S. (1997). On the Efficiency of Visual Selective Attention: Efficient Visual Search Leads to Inefficient Distractor Rejection. *Psychological Science*, 8(5), 395–398. <https://doi.org/10.1111/j.1467-9280.1997.tb00432.x>
- Lewkowicz, D. J., Leo, I., & Simion, F. (2010). Intersensory perception at birth: Newborns match nonhuman primate faces and voices. *Infancy*, 15(1), 46–60. <https://doi.org/10.1111/j.1532-7078.2009.00005.x>
- Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1-6. *Developmental Science*, 17(5), 714–726. <https://doi.org/10.1111/desc.12152>
- Macdonald, J. S. P., & Lavie, N. (2011). Visual perceptual load induces inattentional deafness. *Attention, Perception & Psychophysics*, 73, 1780–1789. <https://doi.org/10.3758/s13414-011-0144-4>
- Mahr, A., & Wentura, D. (2014). Time-compressed spoken word primes crossmodally enhance processing of semantically congruent visual targets. *Attention, Perception, and Psychophysics*, 76(2), 575–590. <https://doi.org/10.3758/s13414-013-0569-z>
- Matusz, P. J., Broadbent, H., Ferrari, J., Forrest, B., Merkley, R., & Scerif, G. (2015). Multi-modal distraction: Insights from children's limited attention. *Cognition*, 136, 156–165. <https://doi.org/10.1016/j.cognition.2014.11.031>
- Matusz, P. J., Dikker, S., Huth, A. G., & Perrodin, C. (2018). Are we ready for real-world neuroscience? *Journal of Cognitive Neuroscience*, (in press).
- Matusz, P. J., & Eimer, M. (2011). Multisensory enhancement of attentional capture in visual search. *Psychonomic Bulletin & Review*, 18(5), 904–909. <https://doi.org/10.3758/s13423-011-0131-8>
- Matusz, P. J., & Eimer, M. (2013). Top-down control of audiovisual search by bimodal search templates. *Psychophysiology*, 50, 996–1009.

<https://doi.org/https://doi.org/10.1111/psyp.12086>

Merkley, R., & Ansari, D. (2016). Why numerical symbols count in the development of mathematical skills: Evidence from brain and behavior. *Current Opinion in Behavioral Sciences*, 10(April), 14–20. <https://doi.org/10.1016/j.cobeha.2016.04.006>

Murray, M. M., & Wallace, M. T. (2012). *The neural bases of multisensory processes*. CRC Press.

Neil, P. A., Chee-Ruiter, C., Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2006). Development of multisensory spatial integration and perception in humans. *Developmental Science*, 9(5), 454–464. <https://doi.org/10.1111/j.1467-7687.2006.00512.x>

Raveh, D., & Lavie, N. (2015). Load-induced inattentional deafness. *Attention, Perception, and Psychophysics*, 77(2), 483–492. <https://doi.org/10.3758/s13414-014-0776-2>

Reynvoet, B., De Smedt, B., & Van den Bussche, E. (2009). Children's representation of symbolic magnitude: The development of the priming distance effect. *Journal of Experimental Child Psychology*, 103(4), 480–489. <https://doi.org/10.1016/j.jecp.2009.01.007>

Santangelo, V., & Spence, C. (2007). Multisensory cues capture spatial attention regardless of perceptual load. *Journal of Experimental Psychology: Human Perception and Performance*, 33(6), 1311–1321. <https://doi.org/10.1037/0096-1523.33.6.1311>

Sarmiento, B. R., Matusz, P. J., Sanabria, D., & Murray, M. M. (2016). Contextual factors multiplex to control multisensory processes. *Human Brain Mapping*, 37, 273–288. <https://doi.org/10.1002/hbm.23030>

Standards and Testing Agency. (2013). Early Years Foundation Stage Profile Handbook. Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/301256/2014_EYFS_handbook.pdf

Sugimoto, F., & Katayama, J. (2017). Increased visual task difficulty enhances attentional capture by both visual and auditory distractor stimuli. *Brain Research*, 1664, 55–62. <https://doi.org/10.1016/j.brainres.2017.03.026>

- Tellinghuisen, D. J., Cohen, A. J., & Cooper, N. J. (2016). Now hear this: Inattentional deafness depends on task relatedness. *Attention, Perception, and Psychophysics*, 78(8), 2527–2546.
<https://doi.org/10.3758/s13414-016-1169-5>
- Tellinghuisen, D. J., & Nowak, E. J. (2003). The inability to ignore auditory distractors as a function of visual task perceptual load. *Perception & Psychophysics*, 65(5), 817–828.
<https://doi.org/http://link.springer.com/article/10.3758/BF03194817#page-1>
- Wagner, J. B., & Johnson, S. C. (2011). An association between understanding cardinality and analog magnitude representations in preschoolers. *Cognition*, 119(1), 10–22.
<https://doi.org/10.1016/j.cognition.2010.11.014>
- Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, 88(3), 638–667.
<https://doi.org/http://www.ncbi.nlm.nih.gov/pubmed/7003641>
- Wu, R., Nako, R., Band, J., Pizzuto, J., Y., G., Scerif, G., & Aslin, R. (2015). Rapid Attentional Selection of Non-native Stimuli despite Perceptual Narrowing. *Journal of Cognitive Neuroscience*, 27, 2299–2307. <https://doi.org/10.1162/jocn>
- Zäske, R., Perlich, M. C., & Schweinberger, S. R. (2016). To hear or not to hear: Voice processing under visual load. *Attention, Perception, and Psychophysics*, 78(5), 1488–1495.
<https://doi.org/10.3758/s13414-016-1119-2>