Vision Research 233 (2025) 108620

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

When children get the gist: The development of rapid scene categorisation☆

Elizabeth A.G. Watson *, Louise Ewing, George L. Malcolm

School of Psychology, University of East Anglia, Norwich, UK

ARTICLE INFO

Keywords:
Scene perception
Cognitive development
Scene categorisation
Gist recognition

ABSTRACT

Research surrounding adult recognition of scene gist is extensive; however, very little is known of its development. Behavioural studies of scene processing tend to broadly support a protracted developmental trajectory, with a quantitative and perhaps also qualitative shift towards more adultlike processing across middle childhood. Here we sought to better understand the very early stages of children's scene processing by targeting gist perception. Children aged 5–10 years categorised backwards-masked scenes presented at very brief durations. We drew inferences about the *processing speed* with which each age group extracted category-diagnostic information by varying presentation durations, and the *quality of information* extracted by varying the level they were prompted to make their judgments (superordinate-level indicative of coarse global information, basic-level indicative of more detailed information). Children across all ages demonstrated a remarkably sophisticated ability to extract scene gist, with 5–6-year-old children performing above chance for scenes presented for as little as 32 ms for both superordinate and basic-level judgements. Categorisation performance also became more efficient with age. Overall, our novel findings indicate that young children possess an impressive ability to process a scene's gist, which is followed by a protracted development towards expertise across middle childhood.

1. Introduction

Our world is comprised of complex and dynamic visual environments, and yet within milliseconds of encountering a novel scene we can piece together the essential features diagnostic of its category. This is referred to as recognising the *gist* of a scene (Greene & Oliva, 2009). People can efficiently recognise scenes presented for even very brief durations, from a single fixation (about 200 ms, see Tatler & Vincent, 2008) to as little as 8 ms (e.g. Furtak et al., 2022). Encoding of scene gist is evident in our neural processes within the very earliest moments of perception, with EEG evidence revealing differences in brain activity within 100 ms of scene onset (Lowe et al., 2018). Since scenes are structured in statistically predictable ways (Kaiser & Cichy, 2021), knowledge of gist also facilitates access to knowledge of layout, including *what* objects to expect and their most likely *locations* (Carrigan et al., 2019; Greene et al., 2016; Võ et al., 2019). The ability to access this information enhances subsequent processing and is critical to acting within and navigating the world with ease.

Despite extensive research with adults, we know relatively little of how gist processing develops. Of those studies that have looked at scene processing more generally in children, emerging evidence supports a particularly protracted development. Quantitative and qualitative differences exist in how children and adults view scenes, with younger children (<7 years) demonstrating longer fixations, and their gaze is also more likely to be captured by visual saliency (Helo et al., 2014). These developmental differences bring into question what information children can extract in a single fixation. Furthermore, research into children's attention towards scenes has suggested the presence of bias towards objects, with 4-year-old children demonstrating difficulties directing attention away from task-irrelevant objects (Darby et al., 2021). Young children have also been found to preferentially represent objects, but not walls, when drawing scenes (Dillon, 2021), and demonstrate a subjective preference for objects, viewing them as more useful and valuable than other scene features (Dillon & Spelke, 2017). This bias towards objects may direct children's attention away from other important diagnostic information within scenes, such as more large-scale structures and overall layout. In the context of scene gist, a strategy that focuses on objects would be disadvantageous at brief durations; diverting limited attentional resources away from the wider spatial distribution of the scene that is important for processing scene images efficiently (Kaiser & Cichy, 2021).

When encountering a new scene, understanding is possible at

☆ This article is part of a special issue entitled: 'Scene Processing' published in Vision Research.

* Corresponding author.
E-mail address: lizzie.watson@uea.ac.uk (E.A.G. Watson).

https://doi.org/10.1016/j.visres.2025.108620
Received 1 October 2024; Received in revised form 15 April 2025; Accepted 7 May 2025
Available online 17 May 2025

different levels of specificity. For example, we can categorise places at a superordinate (e.g., indoors) or basic level (e.g., a kitchen; Tversky & Hemenway, 1983). Conceptualisation at a superordinate level predominantly utilises the coarse, global information available within scenes (Schyns & Oliva, 1994), emerging early in perception (Fei-Fei et al., 2007; Kadar & Ben-Shahar, 2012; Sun et al., 2016). Conceptualisation at the basic-level requires further extraction of detail (Malcolm et al., 2014), suggested by the increased time taken to make basic judgments (Kadar & Ben-Shahar, 2012; Loschky & Larson, 2010). Recent work investigating the development of the scene-selective neural network, a system of regions characterised by their role in scene processing, suggests that the parahippocampal place area (PPA) develops around 5 years of age (see Dilks et al., 2022 for review). This element of the network is particularly sensitive to scene category (Walther et al., 2009) and openness (Kravitz et al., 2011). Regarding basic-level processing, the development of the lateral occipital cortex (LOC), which some suggest supports an object-based channel for scene recognition complimenting the PPA (MacEvoy & Epstein, 2011) continues to develop into adolescence (Nishimura et al., 2015). However, findings surrounding the development of the scene selective network remain notably inconsistent across studies (see Dilks et al., 2023 for review). Here, the theoretical basis behind neuroimaging research for scene categorisation is wanting, i.e., it is not clear how or if these observed patterns of neural development align with behavioural milestones.

The current study is the first of its kind to explore the extraction of scene gist during development. Characterisation of these early perceptual processes constitutes a challenge for researchers even when working with adult participants. One way to investigate these mechanisms is to present backwards-masked scenes at brief durations in a scene gist recognition task. Manipulating the presentation time of scenes offers an accessible way to address the temporal dynamics of scene categorisation, without relying on potentially unreliable measures of response time from children. To investigate the *quality* of information processing, we manipulated the level at which participants were required to make category judgments. We hypothesised that across all durations, performance in scene recognition would improve with age following a gradual transition toward more expert and adultlike processing (Helo et al., 2014). We further predicted basic-level categorisation would be relatively less sophisticated than at the superordinate-level, due to the protracted development of attentional mechanisms (Darby et al., 2021) and object-selective neural structures (Nishimura et al., 2015) that are recruited for the latter, more detail-oriented judgments.

2. Methods

2.1. Participants

The final sample comprised 102 participants: 31 children aged 5–6 years, 35 children aged 7–8 years, and 36 children aged 9–10 years (Table 1). An additional sample of 31 adults was collected (see Supplementary Table 1) but excluded from analyses due to ceiling effects (see Supplementary Table 1). Children were recruited from local science outreach events: 76% at a community science festival, and 24% in a 'pop-up laboratory' at a local library. Adults were recruited and tested from the psychology student population at the host institution. Children received stickers and a certificate, and adults received course credit. The final sample did not include 4 participants (1 × 5–6 years, 1 × 8–9 years & 2 × 9–10 years) who answered *no* to a question at the end of the session asking if we could use their data in our research project. All participants performed above 75% on catch trials included to ensure they were paying close attention to the tasks (Table 1). The study was approved by the School of Psychology Ethics Committee at the University of East Anglia (project ref. ETH2223-0854). Informed consent was provided by the adult participants and parents/legal guardians of all children through an online form (Qualtrics, Provo, UT), and all children provided converging verbal assent.

2.2. Materials and procedures

Visual stimuli consisted of 48 high-definition photographs of natural scenes (see Fig. 1b), taken by the researchers, or from the internet (size: 800 × 600 pixels). These included 24 indoor scenes anticipated to be familiar to children in all age groups (bathroom, bedroom, living room, kitchen) and 24 outdoor scenes (garden, beach, city, playground). 48 unique masks were created from these scene images using MATLAB (v2021B) with Portilla and Simoncelli's (2000) texture synthesis toolbox. Masks were pseudo-randomly assigned to trials and were never assigned to the scene they had been created from. Stimuli were presented using E-Prime 3.0 (Psychology Software Tools, Pittsburgh, PA) on a 52.7 × 29.8 cm monitor (resolution: 1920 × 1080 pixels), with participants sitting at a comfortable distance from the screen.

Participants made 24 superordinate-level and 24 basic-level judgments in two separate blocks. The researcher manually initiated the start of each trial after ensuring children were paying attention. Each trial began with a 500 ms fixation cross, followed by a scene that was presented for one of four randomly assigned durations, 32 ms, 64 ms, 100 ms, 200 ms (balanced within blocks), and then by a visual mask for 50 ms to prevent retinal persistence of the image (see Fig. 1a). Within each block, an additional four oddball "outer-space" catch-trials (highly visually distinct 'space' scenes) were presented for 100 ms to keep participants engaged and serve as an attentional check. Participants had three response options on each trial: two scene labels (one target, one foil) and one for "space". In the superordinate task, the options were always "indoors", "outdoors" and "space". In the basic task, the three options were two basic-level category labels (one of which was correct) and "space". The two basic-level category labels were always from the same superordinate category (i.e., both indoor scenes or both outdoor scenes) and the order of scenes (correct/incorrect) and pairing of scenes were counterbalanced. Participants verbalised their response which was then recorded by the researcher, who sat beside them in a one-on-one interaction and provided effort-based praise and encouragement throughout.

Before each block, five practice trials were completed wherein the presentation durations gradually decreased (1500 ms > 1000 ms > 500 ms > 250 ms) with one additional catch trial ('space' scene) example also included (500 ms). These practice trials used basic-level categories different to the main task (dining room, office, mountain, and lake). Four versions of the experiment were created, in which the order of tasks, and images assigned to each condition were counterbalanced.

3. Results

Considering the exploratory nature of the study, a broad range of durations were applied. Within the superordinate task, children

Table 1
Participant information.

Age Group	*M* Age (*SE*)	Age Range		Gender (*N*)			Neurodevelopmental Condition Status (*N*)			*M* Catch Trial Accuracy (*SE*)	
		Min	Max	Female	Male	Non-Binary	No	Yes	Unsure	Superordinate	Basic
5–6 years	6.08 (0.12)	5.09	6.91	18	13	0	29	1	1	0.96 (0.02)	0.93 (0.02)
7–8 years	8.16 (0.09)	7.16	8.92	17	18	0	34	1	0	0.96 (0.01)	0.96 (0.02)
9–10 years	9.99 (0.08)	9.17	10.90	25	11	0	30	2	4	1.00 (0)	0.96 (0.02)

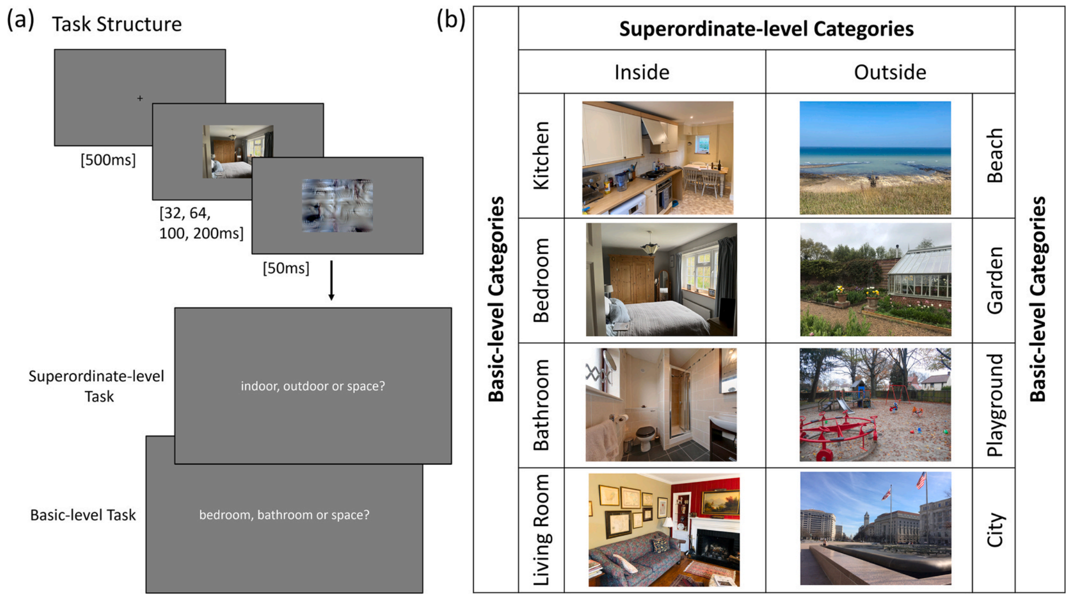

Fig. 1. (a) Overall schematics for task procedure (b) Example stimuli used within the experiment, organised into superordinate- and basic-level categories.

performed exceedingly well at 200 ms, demonstrating a notable ceiling effect (see Supplementary Table 2). For this reason, analyses of the superordinate-level categorisation task data were restricted to the 32, 64 and 100 ms durations. No ceiling effects were observed in the basic-level categorisation task data, thus all four durations were included in the analysis. These analyses were conducted using R Studio (RStudio Team, 2024). For both tasks, one sample t-tests (IBM SPSS Statistics, v29.0.2.0) confirmed that all age groups performed significantly above chance at all durations, *ps* < 0.001 (see Supplementary Table 2). These basic tests confirmed that by 5 years, children are capable of extracting the gist of a scene (both basic and superordinate level) following viewing times as short as 32 ms.

3.1. *Superordinate-level categorisation task*

Firstly, we sought to investigate developmental changes in the *superordinate-level* categorisation task (Fig. 2a). We analysed children's task accuracy scores using a binomial, logistic mixed-effects regression model with duration (32 ms, 64 ms, 100 ms) and age group (5–6 years, 7–8 years and 9–10 years) as predictors and subject number as a random-effects intercept, applying forward difference coding (*lme4,* v1.1–36). Residuals for each model were checked via QQ plots using the *DHARMa* package (v0.4.7). Model BIC scores were compared using Wald Chi-square tests (car package, v3.1–3; See Supplementary Table 3), which was also used to compare the predictors within the final model. Our final model included both duration and age group as predictors (χ^2 (2) $= 15.16, p < 0.001$). Here, we observed a significant effect

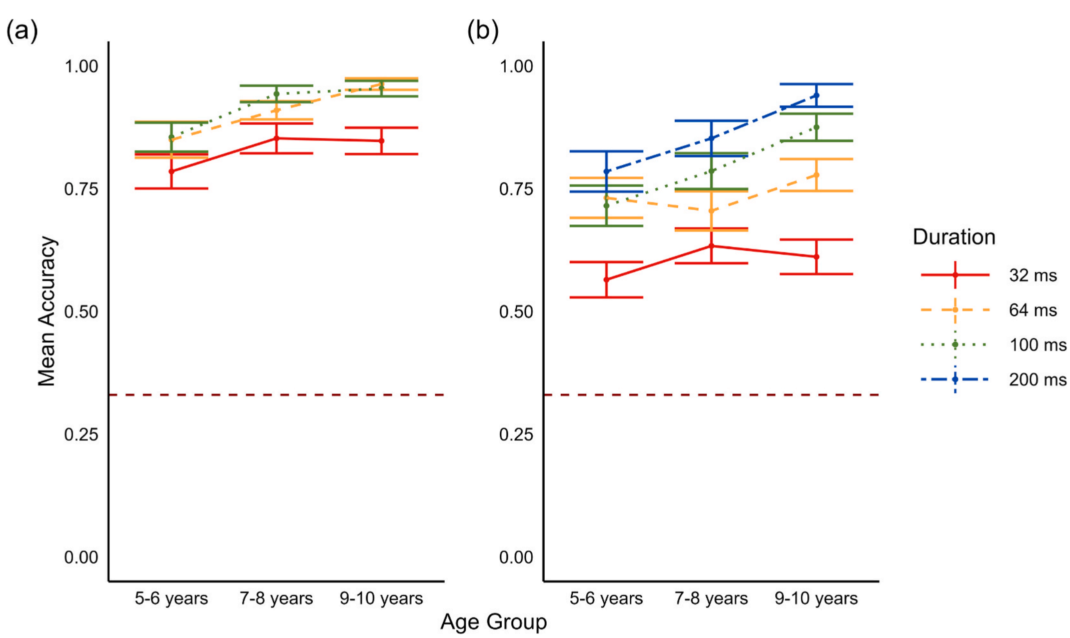

Fig. 2. (a) Superordinate-level Response Accuracy (b) Basic-level Response Accuracy. Mean accuracy scores at each duration, organised by age group with error bars indicating 1 *SE* of the mean. The dashed red line reflects chance at 0.33.

of duration (χ^2 (2) = 30.47, p < 0.001) and age group (χ^2 (2) = 16.42, p < 0.001; see Supplementary Table 4). Including the interaction between duration and age group did not significantly improve the fit of the model (χ^2 (4) = 8.10, p = 0.088), suggesting the developmental effect observed was consistent regardless of participant's access to the scene information (presentation duration).

The pattern of differences underlying these main effects were examined via post-hoc tests using the *emmeans* packages (v1.10.7; test performed on the log scale, odds ratios reported, p values adjusted using Bonferroni method). All post-hoc test results can be found in Table 2. As expected, performance improved with longer durations, with significant improvement observed from 32 to 64, and 32 to 100 ms, but not between 64 and 100 ms. Performance also improved with age across middle childhood, with both 7–8- and 9–10-year-old children performing significantly better than 5–6-year-old children. No further improvement in accuracy was seen between 7–8 and 9–10 years.

3.2. Basic-level categorisation task

A similar approach was taken to analyse the data from the *basic-level categorisation task* (Fig. 2b). Once again we compared accuracy scores using a binomial, logistic mixed-effects regression model, this time including 200 ms within our duration predictor (i.e., 32 ms, 64 ms, 100 ms, and 200 ms) in addition to age group, with subject number as a random intercept, applying forward difference coding. Model BIC scores were compared using Wald Chi-square tests (see Supplementary Table 5). Our final model included duration, age group, and the interaction between duration and age group as predictors (χ^2 (6) = 20.7203, p = 0.002). Here, we observed significant main effects of duration (χ^2 (3) = 122.443, p < 0.001) and age group (χ^2 (2) = 122.443, p < 0.001), in addition to a significant interaction between the two (χ^2 (6) = 19.835, p = 0.003; see Supplementary Table 6).

This significant interaction was explored further with post-hoc tests (performed on the log scale, odds ratios reported, p values adjusted using Bonferroni method). All post-hoc test results can be found in Table 3. Notably, while accuracy improved in the youngest group from 32 to 64/100 ms, they were unable to utilise longer durations more effectively (i.e. no improvement from 64 to 100/200 ms). 7–8-year-old children were also somewhat inconsistent in their improvement across duration, with no improvement in accuracy between the shortest durations (32 and 64 ms) but notable improvement with later durations, between 32 and 100/200 ms, and 64 and 200 ms. By contrast, 9–10-year-old children demonstrated clear improvement between all durations, apart from between 100 and 200 ms.

While each age group's effectiveness in utilising extended durations differed, we were surprised to find no improvement in accuracy with age for stimuli presented at both 32 and 64 ms. Age differences only became notable at 100 ms, with 9–10-year-old children performing significantly better than 5–6-year-olds. At 200 ms, 9–10-year-old children performed better than both the 5–6- and 7–8-year-old groups.

4. Discussion

In the current study, we compared performance in scene gist categorisation across middle childhood (ages 5–10 years). Our main aims were to evaluate children's ability to extract scene gist, varying the presentation time of scenes to infer the processing speed with which children can extract category-based information. We analysed this performance by level of categorisation; first examining children's abilities to extract the broad and coarse information associated with making superordinate-level (inside/outside) judgments, followed by their ability to extract more detailed information diagnostic of the basic-level (e.g. kitchen/bathroom, garden/playground etc.). Previous research has suggested that children undergo significant developmental changes within their perception and understanding of scenes across the targeted age range (Dilks et al., 2022; Helo et al., 2014). Our findings reveal for the first time that young children possess a sophisticated ability to rapidly extract key features of scene images and can accurately categorise scenes at both the superordinate and basic levels even at very short presentation durations. This ability becomes more refined with age, with older children able to extract more information in less time than younger children.

When prompted to categorise scenes at the superordinate level, all children were shown to be proficient. Even 5–6-year-olds performed well above chance from our shortest presentation durations,

Table 2
Post-hoc tests of significant effects within final superordinate model.

Contrast	Odds Ratio	SE	p	Cohen's d
duration 32 ms/duration 64 ms	0.46	0.08	<.001	0.77
duration 32 ms/duration 100 ms	0.41	0.08	<.001	0.90
duration 64 ms/duration 100 ms	0.88	0.18	1.000	0.13
5–6 years/7–8 years	0.51	0.12	0.012	0.67
5–6 years/9–10 years	0.40	0.10	<.001	0.92
7–8 years/9–10 years	0.78	0.19	0.907	0.25

Table 3
Post-hoc tests of significant effects within final basic model.

Contrast	Odds Ratio	SE	P	Cohen's d
Within Age Group				
5–6 years				
duration 32 ms/duration 64 ms	0.45	0.10	0.003	0.81
duration 32 ms/duration 100 ms	0.49	0.11	0.010	0.72
duration 32 ms/duration 200 ms	0.33	0.08	<.001	1.12
duration 64 ms/duration 100 ms	1.09	0.26	1.000	0.09
duration 64 ms/duration 200 ms	0.73	0.18	1.000	0.32
duration 100 ms/duration 200 ms	0.67	0.17	0.634	0.40
7–8 years				
duration 32 ms/duration 64 ms	0.70	0.15	0.627	0.35
duration 32 ms/duration 100 ms	0.44	0.10	0.002	0.81
duration 32 ms/duration 200 ms	0.27	0.07	<.001	1.29
duration 64 ms/duration 100 ms	0.63	0.15	0.289	0.46
duration 64 ms/duration 200 ms	0.39	0.10	0.001	0.94
duration 100 ms/duration 200 ms	0.62	0.16	0.405	0.48
9–10 years				
duration 32 ms/duration 64 ms	0.42	0.09	0.001	0.88
duration 32 ms/duration 100 ms	0.20	0.05	<.001	1.62
duration 32 ms/duration 200 ms	0.09	0.03	<.001	2.46
duration 64 ms/duration 100 ms	0.48	0.13	0.036	0.74
duration 64 ms/duration 200 ms	0.21	0.07	<.001	1.58
duration 100 ms/duration 200 ms	0.43	0.15	0.112	0.84
Between Age Groups				
Duration 32 ms				
5–6 years/7–8 years	0.73	0.19	0.696	0.32
5–6 years/9–10 years	0.81	0.21	1.000	0.22
7–8 years/9–10 years	1.11	0.29	1.000	0.10
Duration 64 ms				
5–6 years/7–8 years	1.15	0.32	1.000	0.14
5–6 years/9–10 years	0.75	0.22	0.944	0.29
7–8 years/9–10 years	0.65	0.18	0.376	0.426
Duration 100 ms				
5–6 years/7–8 years	0.66	0.19	0.464	0.41
5–6 years/9–10 years	0.33	0.10	0.001	1.12
7–8 years/9–10 years	0.49	0.16	0.072	0.71
Duration 200 ms				
5–6 years/7–8 years	0.61	0.19	0.360	0.49
5–6 years/9–10 years	0.21	0.08	<.001	1.56
7–8 years/9–10 years	0.34	0.13	0.016	1.07

demonstrating an impressive ability to extract coarse, spatial information from scenes. Overall, we observed significant improvement in ability to extract the superordinate-level scene gist from 5–6 years to 7–8 and 9–10 years. No further improvement was seen between 7–8 and 9–10 years, suggesting a potential plateau in children's efficiency to extract scene gist at the superordinate level. This pattern of development across middle childhood is consistent with past research suggesting that these ages reflect a key time for scene processing development, e.g., other studies have reported a shift towards adultlike eye movement when viewing scenes (Helo et al., 2014).

Considering the limited number of neuroimaging studies in this area, it is difficult to say if our behavioural findings complement existing knowledge surrounding scene-selective network development. For example, some have proposed that scene categorisation undergoes an accelerated development in childhood, from infancy to 5 years, relative to navigation abilities in line with early presentations of adultlike selectivity to scenes in the PPA (Dilks et al., 2022). Our behavioural findings support this early emergence of categorisation abilities, but also suggest a potentially lengthy course towards scene expertise, mirroring research that has observed later development of the PPA and white matter tracts connecting scene-selective regions, with maturation extending into adolescence (Meissner et al., 2019, 2021).

When categorising scenes at the basic level, children again impressively performed above chance at all durations, and all age groups demonstrated improvement with extended presentation duration. Surprisingly, age differences were only observed when scenes were presented at 100 and 200 ms; the expertise we expected to see in older children was absent when presentation duration was restricted. Our findings further suggested that each age group was able to utilise increases in presentation time to a different extent. Children aged 5–6 years did not benefit from further viewing time after 64 ms, 7–8-year old children did not improve across the shorter durations (32–64 ms), while 9–10-year-old children showed clear improvement across almost all durations. Here, young children's attentional bias towards objects did not prove advantageous (Darby et al., 2021). Instead, differences in children's ability to modulate and distribute attention may have affected their ability to recognise scenes at the basic level. When categorising objects, infants and younger children typically distribute their attention more diffusely as opposed to selectively attending to the most diagnostic features (Best et al., 2013; Huang-Pollock et al., 2011). Additionally, research suggests that children's visual working memory (VWM) capacity improves and becomes adult-like around 6–8 years (see Pailian et al., 2016 for review). These differences in attention and working memory capacity may be further amplified when viewing scenes, which are often more visually complex compared to individual objects.

It is possible that children's processing and understanding of scenes at the basic and superordinate level may profoundly differ. Unfortunately, methodological differences between the design of the tasks administered here mean that we cannot directly compare children's superordinate-level versus basic-level gist recognition. Nevertheless, it is important to consider the possibility that these categorisation abilities could develop following separate trajectories, which would align with past research in adults. Distinctions have been observed between superordinate- and basic-level categorisation, with many studies providing evidence for an early superordinate advantage (Greene & Oliva, 2009; Kadar & Ben-Shahar, 2012; Sun et al., 2016). Different developmental trajectories in childhood could support theories of distinctive but complementary superordinate and basic processing routes within the brain, the latter incorporating more typically object-selective regions. This developmental trajectory supports existing theories that object selective regions such as the LOC work to support scene selective regions including the PPA in scene recognition (Baldassano et al., 2013; Harel et al., 2013; Iordan et al., 2015). For now, our findings suggest investigations of scene categorisation should avoid generalising across these different levels of abstraction if we want to achieve a true understanding of children's scene categorisation abilities.

We have broadly suggested here that the age-related changes in children's performance reflects improved visual processing with age. However, it may alternatively, or additionally, be a result of an improvement in children's *conceptualisations* of scene categories with age. There is limited information available about changes in children's conceptualisation of scenes with age because the literature thus far has primarily focused on objects (see Owen & Barnes, 2021; Poulin-Dubois & Pauen, 2017 for review). Still, such work reports a gradual shift in children's conceptualisation of objects, from a dependency on superficial, physical aspects (4 years; Keil, 2006) towards a more abstract, flexible understanding of taxonomy, as older children (10 years) demonstrate the ability to apply different levels of categorisation at which objects can be understood (Blaye et al., 2006). If the conceptualisation of scenes develops in a similar way with age, then it may be that younger children struggle to consider scenes at multiple levels of abstraction and are thus more rigid in their understanding of scene categories. This flexibility may be particularly relevant to scene processing where the rules surrounding diagnostic features are ambiguous. For example, some scenes belonging to different basic-level categories may share similar objects and/or layouts (e.g. the presence of food in a kitchen/pantry/dining room, or countertops in a kitchen/bathroom/utility room). It is also true that some scenes belonging to the same basic-level category may not share any of the same objects and layouts, yet are recognised the same by their function (Greene et al., 2016; see Malcolm et al., 2016 for review). Highly efficient scene categorisation requires a degree of flexibility and experience that younger children may not yet possess, and this likely becomes particularly relevant when we restrict the amount of information/processing time children have when making such judgements.

5. Conclusions

The current study provides novel evidence for the relatively early emergence, but protracted developmental course, of scene gist recognition in childhood. By 5 years of age, children can accurately extract and categorise scene gist when scenes are presented for as little as 32 ms. Across 5–10 years, children experience a shift towards more expert processing of scene gist with older children extracting more information at short durations. Furthermore, older children showed more sensitivity to viewing time and were better able to utilise longer viewing times to extract information required to make accurate category judgments. While a similar pattern of development was observed across both superordinate- and basic-level scene categorisation, children performed at a more advanced level when making superordinate-level judgments. Taken together, these findings support the view that early scene processing and subsequent categorisation follows a protracted trajectory towards expertise across childhood.

Artificial intelligence

No generative AI or AI-assisted technologies were used during any part of the project.

Ethics

Ethical approval was provided by Ethics Committee of the School of Psychology of the University of East Anglia (project ref. ETH2223-0854).

Data

Anonymised primary and ready-to-analyse data are publicly available at https://osf.io/kbwvn/?view_only=fdc3a5fe1e7d41a9b4680c53b82b936f.

CRediT authorship contribution statement

Elizabeth A.G. Watson: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Louise Ewing:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **George L. Malcolm:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

We would like to thank the Norwich Science Festival and the Norfolk and Norwich Millennium Library for their support. We would also like to thank our brilliant team of experimenters who helped collect data at the festival: Zoe Balaam, Delyth Hughes, Emily Mason, Beatrix Serman, and Kristina Veranic. Finally, thank you to Prerna Aneja and Antje Nuthmann for their advice and support in modelling our data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.visres.2025.108620.

Data availability

Anonymised primary and ready-to-analyse data are publicly available at https://osf.io/kbwvn/?view_only=fdc3a5fe1e7d41a9b4680c53b82b936f.

References

Baldassano, C., Beck, D. M., & Fei-Fei, L. (2013). Differential connectivity within the Parahippocampal Place Area. *NeuroImage, 75*, 228–237. https://doi.org/10.1016/j.neuroimage.2013.02.073

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Best, C. A., Yim, H., & Sloutsky, V. M. (2013). The cost of selective attention in category learning: Developmental differences between adults and infants. *Journal of Experimental Child Psychology, 116*(2), 105–119. https://doi.org/10.1016/j.jecp.2013.05.002

Blaye, A., Bernard-Peyron, V., Paour, J.-L., & Bonthoux, F. (2006). Categorical flexibility in children: Distinguishing response flexibility from conceptual flexibility; the protracted development of taxonomic representations. *European Journal of Developmental Psychology, 3*(2), 163–188. https://doi.org/10.1080/17405620500412267

Carrigan, A. J., Wardle, S. G., & Rich, A. N. (2019). Do target detection and target localization always go together? Extracting information from briefly presented displays. *Attention, Perception & Psychophysics, 81*(8), 2685–2699. https://doi.org/10.3758/s13414-019-01782-9

Darby, K. P., Deng, S. W., Walther, D. B., & Sloutsky, V. M. (2021). The development of attention to objects and scenes: from object-biased to unbiased. *Child Development, 92*(3), 1173–1186. https://doi.org/10.1111/cdev.13469

Dilks, D. D., Jung, Y., & Kamps, F. S. (2023). The development of human cortical scene processing. *Current Directions in Psychological Science, 32*(6), 479–486. https://doi.org/10.1177/09637214231191772

Dilks, D. D., Kamps, F. S., & Persichetti, A. S. (2022). Three cortical scene systems and their development. *Trends in Cognitive Sciences, 26*(2), 117–127. https://doi.org/10.1016/j.tics.2021.11.002

Dillon, M. R. (2021). Rooms without walls: Young children draw objects but not layouts. *Journal of Experimental Psychology: General, 150*(6), 1071–1080. https://doi.org/10.1037/xge0000984

Dillon, M. R., & Spelke, E. S. (2017). Young children's use of surface and object information in drawings of everyday scenes. *Child Development, 88*(5), 1701–1715. https://doi.org/10.1111/cdev.12658

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision, 7*(1), 10. https://doi.org/10.1167/7.1.10

Fox J, Weisberg S (2019). *An R Companion to Applied Regression, Third edition*. Sage, Thousand Oaks CA. https://www.john-fox.ca/Companion/.

Furtak, M., Mudrik, L., & Bola, M. (2022). The forest, the trees, or both? Hierarchy and interactions between gist and object processing during perception of real-world scenes. *Cognition, 221*, Article 104983. https://doi.org/10.1016/j.cognition.2021.104983

Greene, M. R., Baldassano, C., Esteva, A., Beck, D. M., & Fei-Fei, L. (2016). Visual scenes are categorized by function. *Journal of Experimental Psychology. General, 145*(1), 82–94. https://doi.org/10.1037/xge0000129

Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology, 58*(2), 137–176. https://doi.org/10.1016/j.cogpsych.2008.06.001

Harel, A., Kravitz, D. J., & Baker, C. I. (2013). Deconstructing visual scenes in cortex: Gradients of object and spatial layout information. *Cerebral Cortex, 23*(4), 947–957. https://doi.org/10.1093/cercor/bhs091

Hartig, F. (2024). *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level/Mixed) Regression Models. R package version*, (4), 7. https://CRAN.R-roject.org/package=DHARMa.

Helo, A., Pannasch, S., Sirri, L., & Rämä, P. (2014). The maturation of eye movement behavior: Scene viewing characteristics in children and adults. *Vision Research, 103*, 83–91. https://doi.org/10.1016/j.visres.2014.08.006

Huang-Pollock, C. L., Maddox, W. T., & Karalunas, S. L. (2011). Development of implicit and explicit category learning. *Journal of Experimental Child Psychology, 109*(3), 321–335. https://doi.org/10.1016/j.jecp.2011.02.002

IBM Corp. Released 2023. *IBM SPSS Statistics for Windows*, Version 29.0.2.0 Armonk, NY: IBM Corp.

Iordan, M. C., Greene, M. R., Beck, D. M., & Fei-Fei, L. (2015). Basic level category structure emerges gradually across human ventral visual cortex. *Journal of Cognitive Neuroscience, 27*(7), 1427–1446. https://doi.org/10.1162/jocn_a_00790

Kadar, I., & Ben-Shahar, O. (2012). A perceptual paradigm and psychophysical evidence for hierarchy in scene gist processing. *Journal of Vision, 12*(13), 16. https://doi.org/10.1167/12.13.16

Kaiser, D., & Cichy, R. M. (2021). Parts and wholes in scene processing. *Journal of Cognitive Neuroscience, 34*(1), 4–15. https://doi.org/10.1162/jocn_a_01788

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology, 57*, 227–254. https://doi.org/10.1146/annurev.psych.57.102904.190100

Kravitz, D. J., Peng, C. S., & Baker, C. I. (2011). Real-world scene representations in high-level visual cortex: It's the spaces more than the places. *Journal of Neuroscience, 31*(20), 7322–7333. https://doi.org/10.1523/JNEUROSCI.4588-10.2011

Lenth R. (2025). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.10.7-100002, https://rvlenth.github.io/emmeans/.

Loschky, L. C., & Larson, A. M. (2010). The natural/man-made distinction is made before basic-level distinctions in scene gist processing. *Visual Cognition, 18*(4), 513–536. https://doi.org/10.1080/13506280902937606

Lowe, M. X., Rajsic, J., Ferber, S., & Walther, D. B. (2018). Discriminating scene categories from brain activity within 100 milliseconds. *Cortex, 106*, 275–287. https://doi.org/10.1016/j.cortex.2018.06.006

MacEvoy, S. P., & Epstein, R. A. (2011). Constructing scenes from objects in human occipitotemporal cortex. *Nature Neuroscience, 14*(10), 1323–1329. https://doi.org/10.1038/nn.2903

Malcolm, G. L., Groen, I. I. A., & Baker, C. I. (2016). Making sense of real-world scenes. *Trends in Cognitive Sciences, 20*(11), 843–856. https://doi.org/10.1016/j.tics.2016.09.003

Malcolm, G. L., Nuthmann, A., & Schyns, P. G. (2014). Beyond gist: Strategic and incremental information accumulation for scene categorization. *Psychological Science, 25*(5), 1087–1097. https://doi.org/10.1177/0956797614522816

Meissner, T. W., Genç, E., Mädler, B., & Weigelt, S. (2021). Myelin development in visual scene-network tracts beyond late childhood: A multimethod neuroimaging study. *Cortex, 137*, 18–34. https://doi.org/10.1016/j.cortex.2020.12.016

Meissner, T. W., Nordt, M., & Weigelt, S. (2019). Prolonged functional development of the parahippocampal place area and occipital place area. *NeuroImage, 191*, 104–115. https://doi.org/10.1016/j.neuroimage.2019.02.025

Nishimura, M., Scherf, K. S., Zachariou, V., Tarr, M. J., & Behrmann, M. (2015). Size precedes view: Developmental emergence of invariant object representations in lateral occipital complex. *Journal of Cognitive Neuroscience, 27*(3), 474–491. https://doi.org/10.1162/jocn_a_00720

Owen, K., & Barnes, C. (2021). The development of categorization in early childhood: A review. *Early Child Development and Care, 191*(1), 13–20. https://doi.org/10.1080/03004430.2019.1608193

Pailian, H., Libertus, M. E., Feigenson, L., & Halberda, J. (2016). Visual working memory capacity increases between ages 3 and 8 years, controlling for gains in attention, perception, and executive control. *Attention, Perception, & Psychophysics, 78*(6), 1556–1573. https://doi.org/10.3758/s13414-016-1140-5

Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision, 40*(1), 49–70. https://doi.org/10.1023/A:1026553619983

Poulin-Dubois, D., & Pauen, S. (2017). Chapter 26 - The Development of Object Categories: What, When, and How? In H. Cohen & C. Lefebvre (Eds.), *Handbook of Categorization in Cognitive Science (Second Edition)* (pp. 653–671). Elsevier. DOI: 10.1016/B978-0-08-101107-2.00027-0.

Psychology Software Tools, Inc. [E-Prime 3.0]. (2016). Retrieved from https://support.pstnet.com/.

Qualtrics (2023). *Qualtrics*, Provo, Utah, USA: Qualtrics. https://www.qualtrics.com.

RStudio Team (2024). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science, 5*(4), 195–200. https://doi.org/10.1111/j.1467-9280.1994.tb00500.x

Sun, Q., Ren, Y., Zheng, Y., Sun, M., & Zheng, Y. (2016). Superordinate level processing has priority over basic-level processing in scene gist recognition, 2041669516681307 *I-Perception, 7*(6). https://doi.org/10.1177/2041669516681307.

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research, 2*(2), Article 2. https://doi.org/10.16910/jemr.2.2.5

The MathWorks Inc. (2021). *MATLAB version: 9.11 (R2021b)*. Natick, Massachusetts: The MathWorks Inc.

Tversky, B., & Hemenway, K. (1983). Categories of environmental scenes. *Cognitive Psychology, 15*(1), 121–149. https://doi.org/10.1016/0010-0285(83)90006-3

Võ, M.-L.-H., Boettcher, S. E., & Draschkow, D. (2019). Reading scenes: How scene grammar guides attention and aids perception in real-world environments. *Current Opinion in Psychology, 29*, 205–210. https://doi.org/10.1016/j.copsyc.2019.03.009

Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural scene categories revealed in distributed patterns of activity in the human brain. *Journal of Neuroscience, 29*(34), 10573–10581. https://doi.org/10.1523/JNEUROSCI.0559-09.2009