Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

The formation of multisensory object categories:

A developmental perspective

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy to Trinity College Dublin, the University of Dublin.

Eimear M. McKenna, BA. MSc.

Supervised by Professor Fiona N. Newell

School of Psychology and Institute of Neuroscience

Date of conferring January, 2026

Declaration

I, Eimear McKenna, declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work. I confirm that the work has been completed in full compliance with relevant university policies, including the Academic Integrity policy, the Trinity Policy on Good Research Practice, and the guidance on AI and GenAI in Teaching, Learning, Assessment and Research.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

I have read and understood the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at http://www.tcd.ie/calendar. I have also read and understood the guide, and completed the 'Ready Steady Write' Tutorial on avoiding plagiarism, located at https://libguides.tcd.ie/academic-integrity/ready-steady-write

Signed 

Date 31/10/2025

Summary

The perception of objects in everyday life rarely occurs through a single sensory modality. Humans continuously combine and compare information from vision, audition, and touch to identify and categorise objects. These multisensory abilities evolve substantially across development—shifting from early sensitivity to redundant, amodal cues in infancy to increasingly efficient, reliability-based integration strategies in later childhood. This thesis examines how children between four and thirteen years of age use multisensory information to both acquire novel object categories as well as how this information is utilised when children perform categorisation judgements. Following a review of the literature on the development of unisensory and multisensory processing from infancy to adulthood, the empirical work outlined in chapters 2-5 explores how different combinations of visual, auditory, and haptic information contribute to the acquisition and representations of multisensory object categories. Across a series of behavioural experiments, the relative influence of each sensory modality and potential multisensory or cross-modal benefits to object categorisation are systematically assessed.

The studies address several interrelated questions. First, we investigate how audiovisual and contextual cues support the categorisation of familiar objects in children aged 5-13 years (Chapter 2). Next, we examine the acquisition of novel audiovisual dynamic object categories in children aged 5–13 years as compared to adults (Chapter 3), as well as the nature of the subsequent representations in memory. Further experimental work compares visuohaptic and unisensory (visual-only and haptic-only) category learning in children aged 4-13 years, assessing subsequent cross-modal categorisation and generalisation performance (Chapter 4). Finally, using eye-tracking in a visual category learning task, we explore the potential effects of selective attention on learning novel

visuohaptic categories in both children and adults when ones testing modality is known or unknown to be haptic (Chapter 5).

Together, these findings provide new insights into the developmental trajectory of multisensory category learning. The results highlight that while children can exploit mutually redundant multisensory cues, the efficiency and flexibility of this process continue to mature throughout middle childhood. The thesis contributes to a broader understanding of how multisensory experience shapes the formation of conceptual knowledge, offering implications for theories of categorisation, multisensory perceptual learning, and the development of cross-modal perception.

Acknowledgements

Firstly, I want to extend my heartfelt thanks to every school principal, classroom teacher, parent/guardian and the children themselves for taking part in this research. Over the course of my PhD journey I have tested an estimated 1000 children and this could never have happened without the support and interest from the school faculties in the following national schools: Scoil Mhuire Moynalty; Scoil Mhuire Carlanstown; St. Colmcille's S.N.S. Kells; Our Lady of Mercy J.N.S Kells; Carysfort N.S.; Scoil Chaoimhin Marlborough St; St. Christopher's Primary School; St. Vincents Primary School; Kiltale N.S.; John Scottus N.S.; and Central Model Infants' School. This research would not have been possible without the support of Science Foundation Ireland (SFI, grant number 19/FFP/6812), funding training and conferences attendance, and the Department of Psychology at Trinity College Dublin, providing guidance and resources.

To my supervisor Professor Fiona Newell, your insights and guidance have truly been invaluable. At this moment of looking back over the past four years, I want to thank you for the wonderful creative and collaborative atmosphere you have fostered in this research group and the example you have set. Now, to the Multisensory Perception and Cognition Group themselves both past and present: Martina, Alan, Becca, Kate, Beth, Isabella, Nina and Nicole; be it a second pair of hands when testing, or the company shared over a cup of tea, you have all made me laugh on the grim days and been an integral part of this journey.

My friends Ali, Kerry, Lucy, Karen, Natasja, Isabel and Lizzy, you participated when you could, you cheered me on (and up) when you couldn't. Marti my new sister in life and who made this journey what it was through the good and bad times. My siblings Aisling, Meabh and Chloe and finally, Eileen, my mum, my friend and my chief supporter, you taught me to be curious and gave me the confidence to believe I could.

Publications

Publications arising from this thesis (to date):

Newell, F. N., McKenna, E., Seveso, M. A., Devine, I., Alahmad, F., Hirst, R. J., & O'Dowd, A. (2023). Multisensory perception constrains the formation of object categories: A review of evidence from sensory driven and predictive processes on categorical decisions. *Philosophical Transactions of the Royal Society B*, *378*(1886), 20220342. (Chapter 1)

McKenna, E. M., Newell, F.N. (2026). The role of audiovisual object features and scene context on children's ability to categorise familiar objects. *Journal of Experimental Child Psychology, 268*(106506). https://doi.org/10.1016/j.jecp.2026.106506 (Chapter 2)

Table of Contents

List of Figures

List of Tables

Chapter 1

The Development of Multisensory Object Categorisation

Chapter 1

1.0 Introduction: Unisensory Object Perception

Humans are not born with fully developed sensory systems; they go through lifelong changes and adaptations (Gori, 2015). This developmental process and neuronal reorganisation beginning prenatally develops to maturity in early adolescence (Paus, 2005). However, this complex process of sensory maturation is neither ubiquitous among individuals nor the senses themselves. From vision to audition and haptics, the various capabilities of these senses undergo fundamental changes which, over time, lead to adult-like multisensory processors upon the emergence from childhood.

1.1 Visual Perception across development

The development of the human visual system follows a complex but coordinated trajectory involving both peripheral maturation of the eye (Yuodelis & Hendrickson, 1986; Hendrickson et al., 2012; Candy & Banks, 1999) and central reorganization of cortical networks that support visual processing both in primary visual areas (Huttenlocher, 1999; Espinosa & Stryker, 2012) and associative regions within dorsal (Atkinson, 2017; Braddick & Atkinson, 2009) and ventral visual streams (Braddick & Atkinson, 2011; Siu & Murphy, 2018).

These structural and cortical changes support the emergence of visual perceptual abilities from birth through childhood. Newborns exhibit rudimentary selectivity for contour orientation and binocular disparity within the first six weeks of life, suggesting early cortical sensitivity to spatial organization (Braddick et al., 1986; Braddick & Atkinson, 2011). Over the first months of life, acuity and contrast sensitivity improve sharply, paralleling retinal and cortical maturation (Aslin & Banks, 1978; Norcia et al., 1990; Atkinson et al., 2002). Visual acuity improves rapidly across infancy and approaches near-adult levels by five years of age (Brown & Yamamoto, 1986; Atkinson et al., 1974; Neijzen et al., 2025). This limited acuity in infancy is adaptive, allowing infants to focus on proximal and socially salient stimuli such as a caregiver's face (Elman, 1993) while

reducing distraction from background elements (Lickliter, 1990). Orientation selectivity and global form processing also refine progressively throughout early childhood (Siu & Murphy, 2018). Colour discrimination appears by approximately two months, with trichromatic colour vision emerging at approximately three months (Teller et al., 1978). Binocular visual processes such as fusion, depth perception, and stereopsis emerge at around three months and reach adult-like precision by six to seven months (Braddick, 1996). Motion sensitivity develops more slowly: local motion detection emerges during infancy, but global and second-order motion continue to mature into adolescence, typically reaching adult levels between 11 and 14 years (Hadad et al., 2015; Wattam-Bell et al., 2012). This extended period perhaps reflects the complex hierarchical processing demands of motion analysis, requiring precise temporal synchronization and connectivity between cortical regions such as V1, V2, and V5/MT (Ayzenberg & Behrmann, 2024). Collectively, these findings indicate that the human visual system is functional at birth but undergoes substantial refinement across the first decade of life, as the maturation of the eye, cortical architecture, and stream-specific pathways progressively support the emergence of adult-like visual perception.

1.1.1 The Development of Visual Object Recognition.

Visual object recognition is a fundamental cognitive capacity that enables individuals to organise the visual environment into meaningful perceptual units and primarily depends on the ventral visual pathway, which follows a hierarchical structure beginning in early visual areas (V1 and V2) where low-level features such as edges and orientations are extracted. Discrimination of shape and size is evident by approximately 4.5 months (Wilcox, 1999), with the ability to integrate edges and contours emerging by six months (Baker et al., 2008; Taylor et al., 2014). Sensitivity to two-dimensional (2D) component lengths and angles underpins early shape perception (Cohen & Younger, 1984),

while older infants begin to integrate conjunctions and pictorial cues that allow for the perception of three-dimensional (3D) structure (Bhatt & Waters, 1998).

Intermediate visual areas such as V4 are thought to play a central role in shape recognition (Pasupathy & Connor, 2002); with the formation of internal representations of complex objects occurring via shape selective neurons (Pasupathy & Connor, 1999; Wei et al., 2018). As processing progresses through the inferior temporal cortex, neural responses become tuned to increasingly complex visual stimuli. The occipitotemporal cortex (OTC) is particularly implicated in processing 3D shape and supporting perceptual category extraction (Kourtzi et al., 2003). Although low-level features such as texture orientation (Slater et al., 1988) and depth cues (Braddick & Atkinson, 2011) are evident within the first week of life, OTC functionality appears to mature more gradually (Kosakowski et al., 2021). Category-specific representations, such as recognition of animals, typically emerge by 3–4 months of age (Quinn, 2002). Subcortical structures are also critically involved in object recognition; the superior colliculus and maintain reciprocal connections with the ventral visual stream (Gattass et al., 2017). In the mature visual system, they support attentional control and eye movement (Zhou et al., 2016) and appear to contribute to the detection of salient categories such as faces (Almasi & Behrmann, 2021) and animals (Vida & Behrmann, 2017). Infant studies indicate that subcortical integrity is crucial for visual recognition (Pike et al., 1994); with infant object recognition thought to rely predominantly on these subcortical pathways until the ventral cortical regions achieve functional maturity (Johnson, 2001; 2005).

Infant visual perception is said to be conducive to the formation of holistic representations (Ayzenberg & Behrmann, 2022). Children who have experienced visual deprivation at this time and subsequently have their vision restored, reportedly display difficulties in creating a holistic representation for an object and perform poorer at tasks

assessing the perception of global form, though their visual acuity improves to the point of their peers rapidly (Ellemberg et al., 2002). Research on form perception indicates that infants' object judgements rely more on the arrangement of local elements than their individual shapes (Navon, 1977; Cassia et al., 2002). Although infants can discriminate isolated local features, they prioritise global structure when elements are presented as part of a composite form—a phenomenon known as the global precedence effect (Ghim & Eimas, 1988). Immature feedback connections between the lateral occipital complex (LOC) and early visual areas such as V1 likely underlie infants' limitations in global integration (Burkhalter, 1993; Coogan & Van Essen, 1996; Kar et al., 2019). The neural architecture supporting global form perception becomes increasingly integrated through childhood, with feedback connectivity and global processing reaching maturity by adolescence (Káldy & Kovács, 2003; Kovács, 2000; Scherf et al., 2009).

1.1.2 Theories of Visual Object Recognition.

Central to visual object recognition is the capacity to identify an object across varying viewpoints—*viewpoint-invariant object recognition*. This ability requires integration of depth and shape information, as object rotation changes retinal input (Marr & Nishihara, 1978). While adults can recognise unfamiliar objects across orientations (Biederman & Bar, 1999; Tarr & Bülthoff, 1995), infants also display early sensitivity to depth cues (Slater et al., 1991) and binocular disparity (Fox et al., 1980). By three months, infants recognise familiar 3D objects across viewpoints (Mash et al., 2007) and retain these representations for at least 24 hours (Kraebel & Gerhardstein, 2006). By four months, they can use pictorial cues such as shadow and boundaries to infer 3D shape (Bhatt & Waters, 1998) and distinguish between possible and impossible figures (Shuwairi et al., 2007). Multiple theoretical perspectives have been proposed to explain how the visual system achieves object constancy—that is, the capacity to recognise objects across different

viewpoints and contexts. These models can broadly be categorised as **viewpoint-dependent** and **viewpoint-independent** approaches.

Viewpoint-dependent theories posit that object recognition relies on stored representations of multiple views of the same object. Recognition performance thus decreases when an object is presented from an unfamiliar orientation because perception depends on the availability of a matching stored view (Tarr & Bülthoff, 1995). According to this view, repeated exposure to an object from varying perspectives leads to the accumulation of multiple viewpoint-specific representations, facilitating recognition accuracy across orientations. However, developmental findings challenge this account. For example, 3-month-old infants can recognise objects from novel viewpoints that they have never seen before (Kraebel & Gerhardstein, 2006), suggesting that prior exposure to specific perspectives may not be essential for recognition. Such evidence supports the idea that early visual processing may instead rely on more abstract, viewpoint-invariant mechanisms.

Viewpoint-independent theories propose that object recognition relies on representations encoding an object's structural properties irrespective of perspective. Marr's account (1982) is framed within the "tri-level hypothesis" (Peebles & Cooper, 2015), distinguishing computational, algorithmic, and implementational levels of analysis. Marr's theory describes vision as a hierarchical transformation of sensory input through the Primal Sketch, 2.5-D Sketch, and 3-D Model Representation (1982). The Primal Sketch captures local image features such as edges and textures from luminance changes; the 2.5-D Sketch integrates depth and orientation cues to form a viewer-centred representation; and the 3-D Model Representation generates an object-centred, viewpoint-invariant model supporting stable identification. This theory has been critiqued for assuming a purely feedforward hierarchy, with subsequent work highlighting recurrent and interactive cortical

processing (Mareschal & Thomas, 2007; Bennett & Hacker, 2006). Biederman's (1987) Recognition-by-Components (RBC) theory similarly posits structural, viewpoint-invariant representations, suggesting that objects are decomposed into approximately 36 volumetric primitives or "geons" (e.g., cylinders, cones, wedges). Recognition occurs when geons and their spatial relations match stored structural descriptions. Geons are defined by invariant edge and curvature properties; thus, recognition generalises across viewpoints. Nonetheless, the model has been criticised for overlooking contextual and top-down semantic influences (Tarr & Bülthoff, 1995). The Multiple Views theory bridges these perspectives, proposing that the visual system stores several object views along a continuum, allowing flexible interpolation depending on familiarity and task demands. This hybrid approach aligns with neuroimaging evidence that higher-order visual regions, such as the lateral occipital complex (LOC), encode both view-specific and view-invariant features (Kourtzi & Kanwisher, 2001).

Developmental findings support the progressive emergence of these representational mechanisms. Functional near-infrared spectroscopy (fNIRS) studies show that by six months, infants' LOC responses are modulated by object shape (Wilcox et al., 2012), with colour sensitivity emerging by 11 months (Wilcox, 1999; Wilcox & Biondi, 2016). By middle childhood, adult-like LOC activation patterns—supporting viewpoint-invariant recognition of familiar categories such as animals and tools—are evident (Golarai et al., 2007; Dekker et al., 2011). Collectively, these findings suggest that the neural architecture underpinning stable object representations develops gradually from infancy, culminating in the ventral visual stream's mature capacity for invariant yet flexible object recognition.

1.1.3 The development of Visual Object Categorisation.

Evidence suggests that humans use a detailed conceptual taxonomy to aid cognitive efficiency resulting in a novel object being swiftly recognised as belonging to a previously

learned object category. Object categorisation is a fundamental cognitive process that enables humans to manage the vast complexity of perceptual experience by grouping objects according to shared perceptual features or semantic meanings. Through recognising similarities and regularities, categorisation supports inductive reasoning, prediction, and decision-making (Rosch & Mervis, 1975; Murphy, 2002). It underpins conceptual organisation and allows efficient interaction with the environment by forming abstract representations that capture both perceptual and semantic regularities (Ayzenberg & Behrmann, 2024). The process of object categorisation is, therefore, critical to the ability to identify similarities between objects and their common features, classing them into abstract categories due to shared traits, or other fundamental qualities (Sun & Saverese, 2020). This form of object grouping allows rapid deduction of object properties; categorisation due to the basic visual features of an object or its semantic meaning or function is readily observed in adults (Rosch, 1978). A basic level of categorisation would involve a similarity in the global form properties of an object, with local features varying more widely (Rosch et al., 1976). It has long been proposed that the capacity to categorise emerges at an early age, with newborn infants capable of categorising due to a 2D shape's features, for instance closure (Turati et al., 2003). The ability to categorise due to the global structure of a 2D object, such as contour of boundaries, has been observed by 3 months (Quinn, Slater et al., 2001); as well as the capacity to discern between novel and familiar categories for 2D shapes (Bomba & Siqueland, 1983). Regarding real world objects, 2-month-old infants accurately categorised 2D images of objects using a superordinate category of mammals and non-mammals (Quinn & Johnson, 2000; Quinn et al., 1993). The categories formed by infants may additionally be applied to abstract information such as a silhouette (Quinn, Eimas et al., 2001). Research involving 6-month-old infants demonstrated their ability to perform a one-shot categorisation, this refers to the ability to

categorise novel stimuli in accordance with categories they have been exposed to once. This demonstrates the ability to infer category membership due to global object form (Ayzenberg & Lourenco, 2022). Research suggests that from four years of age, the ability to form abstractions to recognise 'family resemblance' between category members to perform category judgments is present (Landau et al., 1998a; Smith et al., 1996).

1.2 Auditory Perception across Development

The auditory system commences development early in gestation, with the cochlea structurally mature by 24–26 weeks (Lavigne-Rebillard & Pujol, 1988) and the cochlear nerve and brainstem nuclei near completion, although myelination continues through the perinatal period (27–29 weeks; Moore & Linthicum, 2001). Myelination proceeds along the auditory pathway from the cochlea through the brainstem toward the medial geniculate body (Yakovlev & Lecours, 1967), this enables sound responsiveness by the third trimester (Birnholz & Benacerraf, 1983; Krumholz et al., 1985). Although the cochlea assumes an adult-like structure by the end of the second trimester, research shows that fine-tuning for high frequencies above 6000 Hz continues into late gestation (Eggermont et al., 1996), with higher-frequency sensitivity maturing by about six months after birth (Folsom & Wynne, 1987; Abdala & Folsom, 1995). The perinatal period also marks maturation of the olivocochlear efferent system, characterised by neuronal enlargement and axonal refinement (Moore et al., 1998; Moore & Linthicum, 2007). Concurrently, the brainstem undergoes rapid differentiation, with auditory neurons enlarging (Gandolfi et al., 1981; Nara et al., 1996) and branching extensively by birth (Moore et al., 1998). Cortical development during the perinatal period leads to differentiation of primary and secondary auditory cortices within the temporal lobe, laying the foundation for complex sound analysis. By birth, infants thus possess mature peripheral and brainstem circuitry but an immature cortex (Werner, 2019), providing a foundation for more complex auditory

processing. As a result, newborns demonstrate limited temporal, frequency, and intensity resolution, which gradually refine over infancy and childhood.

Newborns show attentional bias toward human speech (Trehub, 1973; Vouloumanos & Werker, 2004) and can discriminate a wide range of phonemic and syllabic contrasts (Trehub, 1973; Eilers & Minifie, 1975; Eimas, 1975; Aslin et al., 1981; Jusczyk & Thompson, 1978; Bertoncini et al., 1988), including non-native contrasts (Trehub, 1976; Werker et al., 1981). Temporal discrimination abilities, including temporal resolution and duration discrimination improve gradually across development, in infancy gap detection thresholds range from 40–50 ms (Werner et al., 1992; 1994), reaching adult levels by four to five years (Trehub et al., 1995; Wightman et al., 1989). Duration discrimination remains protracted (Elfenbein et al., 1993; Morrongiello & Trehub, 1987) with frequency discrimination following a similar pattern: sensitivity to high frequencies (~4000 Hz) approaches maturity by six months, but low-frequency discrimination (500–1000 Hz) continues into late childhood (Olsho, 1984; Olsho et al., 1987; 1988). Intensity discrimination is coarse in infancy: six- to nine-month-olds require 6–8 dB changes compared with adults' 1–2 dB thresholds (Sinnott & Aslin, 1985), reaching maturity around five to six years (Jensen & Neff, 1993; Maxon & Hochberg, 1982). Although cochlear and evoked-potential measures appear adult-like (Abdala, 2000; Gorga, 1989), psychophysical thresholds mature slowly. Loudness perception, integrating duration, bandwidth, and frequency, appears relatively adult-like by infancy (Leibold & Werner, 2002; Hellman et al., 1997; Collins & Gesheider, 1989). Threshold sensitivity improves postnatally as the middle and inner ear mature (Ehret, 1997; Lecanuet & Schaal, 1996). However, sensitivity to complex acoustic environments continues maturing beyond early childhood (Litovsky, 1997). Higher-order auditory skills such as speech perception in noise and auditory spatial judgments are highly experience-dependent, with speech-in-noise discrimination maturing

in late childhood (Johnson, 2000) and audio-space bisection reaching adult-like levels around nine years (Gori, Sandini et al., 2012). Infants and young children thus remain "inefficient listeners," showing reduced sensitivity and greater variability due to immaturity in auditory, attentional, and memory systems (Allen & Wightman, 1994; Bargones et al., 1995; Hartley et al., 2000; Werner & Boike, 2001). Collectively, these findings highlight that auditory perception emerges through a prolonged interplay between peripheral maturation, cortical specialisation, and experience-driven refinement across infancy and childhood.

1.2.1 The development of Auditory Object Recognition.

Conceptually, an "auditory object" is the outcome of grouping spectrotemporal regularities into coherent streams (e.g., a voice, a musical instrument), a process long formalized in auditory scene analysis and now framed neurally as the transformation from acoustic to object-based codes in auditory cortex (Bregman, 1994; Bizley & Cohen, 2013). Evidence across primates and humans supports partially segregated cortical pathways: a ventral "what" stream for object identity and a dorsal "where/how" stream for spatial and sensorimotor aspects, with converging support from neurophysiology and fMRI (Rauschecker & Tian, 2000; Belin & Zatorre, 2000; Arnott et al., 2004). Within this framework human neuroimaging shows that regions along superior temporal cortex are parametrically sensitive to vocal harmonic content in natural and synthetic sounds (Lewis et al., 2009; Talkington et al., 2013). Computational and physiological work further implicates temporal coherence—the synchronous modulation of a source's features—as a binding principle that yields object representations downstream from early feature maps (Shamma et al., 2011; Teki et al., 2013; Lu et al., 2017). Developmentally, newborns and young infants already exhibit rudiments of object formation: mismatch-style paradigms reveal organization of the auditory world at birth (Winkler et al., 2003), and both newborns and 4-month-olds use harmonicity/mistuning to segregate concurrent sources, including

perceiving a mistuned partial as a separate object (Bendixen et al., 2015; Folland et al., 2012; Smith et al., 2017). Complementing these behavioral and ERP findings, infant neuroimaging demonstrates voice-selective responses by 3–7 months and voice categorization by ~4 months, indicating early specialization within temporal cortex for biologically salient auditory objects (Blasi et al., 2011; Calce et al., 2024). Although elements of object coding appear even in early-stage auditory areas (e.g., repetition suppression for sound-source meaning), mature efficiency in noisy, reverberant contexts continues to improve through childhood, consistent with prolonged refinement of attention and higher-order cortical networks that support stable, invariant object recognition (Da Costa et al., 2015; Bizley & Cohen, 2013; Calcus, 2024).

1.2.2 The development of Auditory Object Categorisation.

Auditory object categorisation, the mapping of sound sources onto abstract classes such as living versus non-living and vocal versus non-vocal, appears to emerge from early infancy through acoustic grouping mechanisms (Werner, 2019). Nevertheless, this process exhibits a protracted developmental course shaped by cortical specialisation and executive control. Early neuropsychology and neuroimaging studies first established that conceptual knowledge is organised along animate–inanimate lines, with partially dissociable neural substrates that generalise to audition (e.g., animals/tools; Martin et al., 1996; Moore & Price, 1999). In the auditory domain, fMRI research indicates that sounds produced by living versus non-living sources are associated with activity in distinct cortical networks that supports category-level organisation of action and environmental sounds (Engel et al., 2009). A key acoustic–semantic cue that scaffolds category formation is harmonicity: vocalisations tend to carry higher harmonics-to-noise ratios (HNR), and parametrically increasing harmonic content strengthens responses in temporal cortex, aiding categorisation of vocal versus non-vocal objects (Lewis et al., 2009). Research grounded in embodiment propose that repeated audio-motor experiences with conspecific actions and

vocalisations enrich category representations, linking perception with sensorimotor systems (Barsalou et al., 2003; Barsalou, 2008). Developmentally, children's categorisation of everyday environmental sounds shows improvements in strategy use and accuracy across school age (Berland et al., 2015), and rule-based auditory category learning continues to mature into adolescence and early adulthood, with developmental gains tightly coupled to executive flexibility (Reetzke et al., 2016). Together, converging evidence supports a model in which acoustic regularities (e.g., harmonicity) seed early category structure, while experience-dependent cortical specialisation and developing executive functions yield increasingly efficient, abstract auditory object categories across childhood.

1.3 Haptic Perception across Development

Haptic perception—often referred to as active touch—arises from the activation of mechanoreceptors in the skin, muscles, tendons, and joints during exploration of objects through manual movement in space (Klatzky et al., 1985). Through this system, individuals can recognise objects as well as specific features, such as size, shape, weight, material properties such as texture, or the position of object parts. A defining aspect of the haptic modality is its reliance on direct physical contact. The "tactile perceptual field," representing the skin area in contact with a stimulus, is spatially constrained—limited to the surface area of the hands and the reach of the arms, both of which are in periods of protracted growth over the course of the first years of life (Bleyenheuft et al., 2009). Haptic perception, as it is limited by the perceptual field, requires the movement of both hands and arms to explore an object's spatial characteristics (Gentaz et al., 2008). From infancy the capability to haptically explore (Streri & Gentaz, 2004) and manipulate (Streri et al., 2000; Striano & Bushnell, 2005) objects as well as perform object discrimination tasks (Streri & Gentaz, 2003) emerge. Nonetheless, children's physical form, hands and corresponding neurophysiology are experiencing persistent and swift alterations across development. For example, the ability to haptically discriminate size (Gori, Squeri, et al., 2012) and

orientation is evident in late childhood. Moreover, haptic object recognition (Morrongiello et al., 1994) and manipulation (Rentschler et al., 2004), as well as sensorimotor temporal order sensitivity undergo recalibration (Vercillo et al., 2017), improving between the ages of five to seven years (Withagen et al., 2012) and reaching maturity around the age of eight.

Regarding the use of haptics in developmental object perception research, it has been found that performance across age groups can vary depending on the nature of the object feature being explored e.g., shape versus texture (Gliner, 1967; Kaur et al., 2022). Therefore, although haptic perception generally improves to adult-like levels of performance during childhood, the age at which haptic maturation is acquired varies across haptic cues. For example, children's ability to haptically discriminate texture and shape improves rapidly across the school years, but at different rates, suggesting feature specific developmental trajectories (Kaur et al., 2022). Furthermore, in adults the processing of haptic input differs from the visual or auditory system, with equal weighting being given to the different object cues when haptically exploring an object. Regarding children, this proposed activity is both more diffuse as well as less intense, which mirrors the findings of Gori et al. (2008) that spatial discrimination is not supported by the associated neural mechanisms until age eight. Nevertheless, at a young age, haptics can provide valuable insights into object perception, though cortical regions supporting haptic object processing, including the primary and secondary somatosensory cortices and posterior parietal areas, undergo progressive refinement with experience, enabling adult-like integration of tactile cues into coherent object representations (Lacey et al., 2010). Haptic perception matures through a prolonged and experience-dependent process driven by concurrent physical, neural, and motor development, reaching its functional efficiency only toward late childhood.

1.3.1 The development of Haptic object Recognition.

Haptic object recognition emerges early in human development but undergoes a protracted course of refinement throughout childhood. Evidence from neonatal research demonstrates that even newborns possess haptic recognition abilities: they can discriminate between shapes and transfer shape information across hands, suggesting the presence of early somatosensory representations (Streri, 2005; Streri & Gentaz, 2003). At two to three months, infants begin to explore objects with more deliberate grasping and mouthing behaviours that allow them to extract information about texture and form (Striano & Bushnell, 2005; Sann & Streri, 2007). Moreover, by four to six months, infants can recognize previously encountered objects through touch alone and distinguish novel objects, providing compelling evidence that haptic recognition processes are functional well before the onset of coordinated visual–manual exploration (Hatwell et al., 2003; Lejeune et al., 2010). Despite these early competencies, haptic object recognition remains less efficient than in adulthood, as young children rely predominantly on local or textural cues rather than global or spatial features (Tortelli et al., 2023).

Through preschool and early school years, the efficiency of haptic recognition increases substantially as exploratory hand movements and motor coordination mature. Research on exploratory procedures—such as contour following, lateral motion, and enclosure—demonstrates that these movements become more targeted with age, aligning with the diagnostic features most relevant to the recognition task (Lederman & Klatzky, 1987; Kalagher & Jones, 2011a). Morrongiello, Humphrey et al. (1994) showed that children as young as five years can identify familiar objects by touch alone, but accuracy and response speed improve markedly between ages five and ten, coinciding with increased efficiency in feature extraction and somatosensory–parietal cortical maturation. Object manipulation and the ability to encode three-dimensional structural configurations also develop gradually, with studies indicating that configural coding and object-centered spatial

mapping continue to improve across the primary school years (Rentschler et al., 2004). In parallel, sensorimotor timing mechanisms that underpin stable haptic representations, particularly the temporal alignment between tactile feedback and movement, remain immature through late childhood, as older children (8–11 years) still show limited sensory–motor temporal recalibration after delayed feedback (Gori et al., 2013; Vercillo et al., 2014).

A key feature of this developmental trajectory is the increasing precision with which children use touch to discriminate object properties. Touch provides particularly reliable cues for object size and material, and during early development it often dominates other sensory inputs in size estimation tasks, underscoring the robustness of haptic information for spatial and metric judgments (Gori et al., 2008; Petrini et al., 2014). More recent findings also reveal that individual differences in haptic shape-based recognition are predictive of general object recognition performance, suggesting shared higher-order mechanisms for extracting invariant structural information across objects, regardless of modality (Chow et al., 2022). Neuroimaging evidence in adults indicates that regions within the somatosensory and posterior parietal cortices support the transformation of tactile inputs into abstract object representations, with increasing functional specialization observed through development as tactile experience accumulates (Lacey et al., 2010). Collectively, these studies converge on a view that haptic object recognition is functional from infancy but continues to evolve as exploratory procedures, spatial encoding, and temporal calibration mechanisms mature, reaching full efficiency only toward late childhood.

1.3.2 The development of Haptic object Categorisation.

The ability to group objects according to haptically perceived features such as shape, texture, size, and weight—emerges during infancy and follows a protracted, feature-specific developmental trajectory extending into middle childhood. In the first year of life, infants demonstrate early sensitivity to both material and geometric properties through

active touch. By approximately three months, infants can discriminate between shapes by holding and mouthing objects (Striano & Bushnell, 2005) with discrimination of textures reported in approximately seven-month-old infants (Stack & Tsonis, 1999); by six to eight months, they retain haptically derived information across brief delays, suggesting the formation of primitive categorical representations based on tactile similarity (Catherwood, 1993; Streri & Gentaz, 2004). These early categorizations rely heavily on global, salient features such as overall contour and surface roughness rather than fine-grained spatial detail (Striano & Bushnell, 2005), consistent with the immaturity of tactile spatial resolution and exploratory hand control in infancy (Bushnell & Boudreau, 1998; Kalagher & Jones, 2011a). During the preschool years, between approximately three and five years of age, children begin to organize objects into groups according to more specific tactile dimensions. At this stage, texture and material properties dominate categorisation judgments: children typically group objects by how they "feel" (e.g., smooth, soft, rough) rather than by their shape or size (Schwarzer et al., 1999; Kalagher & Jones, 2011a).

Studies using haptic sorting tasks show that children below six years tend to rely on a single salient cue, with texture producing the most consistent classification performance, whereas shape and weight cues are often ignored unless explicitly highlighted (Kalagher & Jones, 2011a; Overvliet et al., 2024). This preference for material-based classification reflects both the perceptual salience of surface cues and immature integration of spatial information from kinaesthetic feedback. Between six and nine years of age, a clear shift occurs toward increased use of spatial dimensions such as shape and size in haptic categorisation. Experimental evidence indicates that by about seven to eight years, children begin to use shape cues more consistently when sorting unfamiliar objects, and their accuracy in distinguishing shape categories improves substantially (Overvliet et al., 2024; AlAhmed et al., 2023).

Infants are sensitive to gross heaviness differences during action and by 9–12 months they can scale their lifting forces to object weight after experience, showing early awareness of weight through action (Mash, 2007; Paulus & Hauf, 2011). However, reliable, property-specific haptic classification and the anticipatory force control needed for weight-based judgments show a protracted trajectory, with marked improvements across early school age (approximately 7–9 years) as exploratory strategies, force control, and proprioceptive precision become more adult-like (Forssberg et al., 1995; King et al., 2012). During this period, children's haptic categorization strategies become increasingly organized around specific tactile dimensions, including weight, rather than global impressions (Bushnell & Boudreau, 1993; Kalagher & Jones, 2011a). Similarly, the capacity to categorize objects by size progresses steadily; infants as young as six months can detect relative size differences, but consistent, rule-based size categorisation is not observed until middle childhood (Gori, Squeri, et al., 2012). By late childhood (around nine to twelve years), haptic categorisation becomes multidimensional and flexible. Children now integrate information from multiple tactile features—combining shape, size, texture, and weight—to form abstract category representations like those of adults (Kalagher & Jones, 2011a; 2011b). This developmental transition is supported by increased efficiency in exploratory procedures, refined tactile spatial acuity, and maturing attentional control over diagnostic cues. Collectively, these findings indicate that haptic object categorisation begins with broad, texture-dominated groupings in infancy and preschool years, then shifts toward spatially structured, shape-based and multidimensional categorisation in later childhood as both perceptual and cognitive mechanisms reach maturity.

2.0 Multisensory Integration Across Development

2.1 The nature of bimodal interactions

Perceiving the world as coherent depends on the brain's ability to combine information from multiple sensory systems. Objects and events in the environment

typically stimulate more than one sense — for instance, when one can seeing and hearing your co-worker typing on a keyboard. The process by which these separate sensory inputs are combined into a single perceptual representation is known as multisensory integration (Stein & Meredith, 1993; Stein, et al., 2020). Through integration, redundant sensory cues that signal the same property of an object (e.g., shape, rhythm) can be merged to enhance perceptual salience, improve accuracy, and support adaptive behaviour (Bahrick et al., 2004; Lewkowicz & Flom, 2014).

To facilitate appropriate integration of sensory information, the brain must determine which sensory inputs arise from the same external source and should therefore be bound together. The brain achieves this through causal inference (Noppeney, 2021), computing whether signals are sufficiently aligned in space, time, and meaning to warrant integration (Körding et al., 2007; Wallace et al., 2020). Integration thus operates alongside a complementary process of segregation; whereby unrelated signals are kept separate to avoid perceptual confusion. Neurophysiologically, integration was first characterised by superadditive neural responses—instances where multisensory input evokes stronger neuronal activity than the sum of the corresponding unisensory responses (Stein & Meredith, 1993; Wallace & Stein, 1997). Importantly, multisensory processing is not merely a perceptual phenomenon but a product of interacting perceptual, cognitive, and motor systems. The capacity to align sensory input across modalities supports motor coordination (Coats et al., 2015), attention allocation (Santangelo & Spence, 2007), and higher-level cognitive processes such as categorisation (Calvert et al., 2004). As such, the development of integration is intertwined with broader cognitive maturation and experiential learning (Bremner et al., 2008).

To accurately chart the developmental trajectory of multisensory processing, which at the broadest level refers to perceptual or neural activity which incorporates information

from more than one sensory modality (Calvert et al., 2004; Stein et al., 2020) clarification is required to define several closely related but functionally distinct processes it encompasses: crossmodal perception, cross-sensory calibration and multisensory integration. Crossmodal perception refers to the ability to transfer or match information across sensory modalities—for instance, recognising an object visually after having only touched it (Meltzoff & Borton, 1979; Streri & Gentaz, 2004), hereafter referred to as crossmodal transfer. This ability emerges early in infancy, reflecting the brain's sensitivity to *amodal properties* of stimuli such as shape, rhythm (Lewkowicz, 2000), or intensity that can be conveyed through multiple senses (Bahrick & Lickliter, 2000). However, crossmodal processing is not restricted to stepwise transfer between senses; it also includes more immediate interactions such as the recruitment or modulation of one sensory system during processing in another (Lacey & Sathian, 2014; Amedi et al., 2001).

Cross-sensory calibration describes the process by which one sensory modality fine-tunes another during development (Ernst & Banks, 2002; Gori, 2015). Unlike integration, calibration does not fuse inputs to produce a unified percept; rather, one sense serves as a *reference* for adjusting another's perceptual accuracy. For example, vision may calibrate haptic perception of orientation, while touch calibrates visual size perception (Gori et al., 2008). Multisensory integration, in contrast, involves the simultaneous use of information from multiple senses to generate a single, more reliable estimate of an object's properties or location. This process improves perceptual precision through weighted averaging based on sensory reliability, a principle formalised in Bayesian cue integration models (Ernst & Bülthoff, 2004; Ma et al., 2006). Clarifying these distinctions resolves a common source of confusion in developmental research. Apparent failures of integration in young children may not reflect an absence of multisensory ability per se, but rather a

period in which cross-sensory calibration predominates as the sensory systems mutually refine one another (Gori, 2015).

Recent research further distinguishes multisensory integration from crossmodal recalibration, two processes that often co-occur but serve distinct functions. Integration refers to the instantaneous combination of sensory cues to produce a unified percept, whereas recalibration describes the longer-term adjustment of one modality's representation based on discrepancies with another (Rohlf et al., 2020). Using a child-adapted ventriloquist paradigm, Rohlf et al. (2020) showed that children aged 6–7 years exhibited immediate ventriloquist aftereffects—short-term shifts in auditory localisation toward visual cues—while older children (8–9 years) demonstrated both integration and cumulative recalibration. This suggests that integration mechanisms precede and may even enable recalibration, providing the multisensory correspondence necessary for aligning sensory maps. The dissociation also helps clarify developmental findings: younger children may show *transient* integration effects without sustained recalibration because their systems have not yet stabilised modality-specific reference frames. Over time, recalibration mechanisms consolidate these relationships, leading to durable perceptual alignment across senses.

Empirical research over the past four decades reveals that the capacity to relate information across modalities begins remarkably early. Infants as young as one month can recognise a pacifier they have sucked when later presented visually (Meltzoff & Borton, 1979), and by two months can visually identify shapes previously explored through touch (Sann & Streri, 2007). Likewise, infants detect equivalence between a person's facial movements and vocalisations within the first four months of life (Dodd, 1979; Patterson & Werker, 2002). These findings demonstrate that even preverbal infants can extract *amodal relations*—shared temporal or spatial structures—across senses (Bahrick, 2001;

Lewkowicz, 1992). However, early multisensory abilities are not fully mature. Young infants' integration often depends on redundant, temporally synchronous, and spatially co-located cues. For example, they prefer looking at audiovisual events that are congruent (e.g., a bouncing ball accompanied by a synchronous sound) over incongruent ones, but fail to match asynchronous pairings (Lewkowicz, 1996; Bahrick & Lickliter, 2000). As development proceeds, this broad sensitivity narrows to reflect experience-dependent tuning—an example of perceptual narrowing—whereby infants become more selective for the temporal and spatial contingencies common in their environment (Lewkowicz & Ghazanfar, 2009).

Developmental theorists propose that multisensory perception progresses through three broad phases: immature, broadly tuned, and narrowly tuned stages (Lewkowicz, 2014). During the immature phase, perception is dominated by low-level stimulus properties such as synchrony and proximity, enabling coarse matching across modalities. The broadly tuned stage marks the onset of learning-based refinement, as infants accumulate experience linking specific sensory combinations to environmental regularities. Finally, in the narrowly tuned phase, children selectively integrate familiar, causally coherent sensory combinations while disregarding irrelevant or conflicting signals. These transitions are shaped by both bottom-up factors, including the maturation of sensory systems and neural transmission, and top-down influences such as attention, prior knowledge, and goals (Bahrick et al., 2004; Hillock-Dunn & Wallace, 2012). The interplay between these factors explains the heterogeneity in developmental timing across individuals and modalities: while temporal binding precision may reach adult-like levels only in adolescence, some spatial localisation abilities emerge much earlier (Neil et al., 2006; Barutchu et al., 2009).

Neurophysiological studies underscore that the ability to integrate sensory information is experience-dependent and develops through postnatal interactions between the senses and the environment. Pioneering work by Stein et al. (1999) demonstrated that neurons in the superior colliculus of newborn cats initially respond to only one modality and fail to exhibit multisensory enhancement. Following normal sensory experience, these neurons acquire the capacity to respond to combined visual and auditory inputs in a superadditive fashion (Stein et al., 1973; Wallace & Stein, 1997). When deprived of such multisensory experience — for instance, through rearing in darkness — animals fail to develop this integrative capacity (Wallace et al., 2004). These findings provide compelling evidence that multisensory neurons rely on cross-modal experience to tune their responses, linking developmental physiology with behavioural maturation. More recent computational and neurophysiological models (e.g., Cuppini et al., 2011; Shaikh, 2022; Wallace et al., 2020) demonstrate that with increasing sensory experience, neural computations shift from competitive to cooperative processing, allowing congruent signals to amplify each other rather than compete.

Taken together, early multisensory abilities reflect a foundation for later perceptual expertise rather than evidence of mature integration. The transition from broad crossmodal sensitivity to selective integration depends on both neural maturation and experience-based calibration. These processes jointly establish the architecture for efficient multisensory processing across spatial and temporal domains. In the following section, I examine how spatial and temporal factors—particularly synchrony, contiguity, and modality appropriateness—shape the developmental trajectory of multisensory integration, and how cross-sensory calibration contributes to these patterns.

2.2. Spatial and Temporal Factors in Multisensory Integration Across Development

The developmental trajectory of multisensory integration is shaped by how the brain learns to coordinate temporal and spatial correspondences between sensory inputs.

Temporal synchrony and spatial congruence are two principal cues that determine whether signals are interpreted as arising from a common cause and therefore integrated (Stein & Meredith, 1993; Wallace & Stein, 1997). Across development, sensitivity to these cues becomes progressively refined as neural networks mature and as perceptual experience constrains which combinations of inputs are likely to be meaningful.

2.2.1 Temporal Synchrony.

Temporal synchrony — the coincidence of sensory events in time — is a critical determinant of multisensory binding. Infants are initially sensitive to a broad range of temporal asynchronies between auditory and visual inputs. Over time, this tolerance narrows, yielding more precise alignment of perceptual timing. Behaviourally, this is reflected in the multisensory temporal binding window (TBW), the interval within which signals from different modalities are perceived as simultaneous and integrated into a unified percept (Lewkowicz, 1996; Vroomen & Keetels, 2010). Lewkowicz and Flom (2014) proposed that the narrowing of this window reflects the combined influence of perceptual experience and neural maturation. During early infancy, broad temporal tuning allows infants to link loosely correlated sensory events, supporting learning about the statistical regularities of the environment. As sensory systems mature and experience accumulates, perceptual precision increases and the TBW contracts. For example, for non-speech audiovisual events, the TBW decreases from approximately 450 ms in infancy to around 150–200 ms in adulthood (Lewkowicz, 1996).

Neurodevelopmentally, this refinement coincides with improvements in myelination, synaptic pruning, and increased transmission efficiency across sensory cortices (Paus et al., 2001; Giedd et al., 1999). These neural changes shorten conduction delays and enhance the temporal fidelity of multisensory correspondence. Behavioural studies parallel these findings: Stevenson et al. (2018) demonstrated that temporal acuity follows a U-shaped trajectory across the lifespan, with precision improving through

childhood and peaking in adolescence. Similarly, Hillock-Dunn and Wallace (2012) observed that children gradually become more accurate at detecting audiovisual synchrony, with adult-like performance emerging only in late childhood. The McGurk illusion—where conflicting auditory and visual speech cues (e.g., hearing “ba” while seeing “ga”) lead to the perception of a fused phoneme (“da”)—provides a classic measure of audiovisual temporal integration. This illusion has been observed in infants as young as four months (Burnham & Dodd, 2004), yet its strength and consistency increase with age, reflecting growing sensitivity to temporal and phonetic congruence. Similarly, visuotactile simultaneity detection reaches adult levels only by around 9–11 years of age (Chen et al., 2018), suggesting that multisensory temporal integration is a protracted developmental process reliant on neural and experiential refinement.

Overall, these findings indicate that multisensory integration in the temporal domain develops from broad tolerance and redundancy detection in infancy toward selective, high-precision binding in later childhood. The narrowing of the temporal binding window thus represents a key index of multisensory maturation, enabling children to align complex, time-varying sensory streams in a way that supports speech, motion perception, and coordinated action.

2.2.2 Spatial Correspondence.

Spatial alignment between sensory cues can also influence the degree to which these cues are integrated. Infants begin by responding reflexively to multisensory events that occur in the same general region of space, but their ability to localise and integrate inputs across modalities sharpens considerably across the first ten months of life (Neil et al., 2006). A key principle governing spatial (and temporal) integration is the modality appropriateness hypothesis, which posits that the modality most accurate or reliable for a given task dominates perception (Welch & Warren, 1980). For spatial localisation, vision typically provides the most precise information, while audition offers superior temporal

resolution. In developmental contexts, this principle interacts with the maturation of each sensory system and with the cognitive capacity to evaluate sensory reliability.

Infants display an early form of auditory dominance, likely due to the relatively earlier maturation of the auditory system and the transient nature of auditory stimuli (Robinson & Sloutsky, 2004a). When presented with audiovisual pairings, infants and preschoolers often encode auditory information at the expense of visual features, even when they can recognise those features in unimodal conditions (Sloutsky & Napolitano, 2003). However, this bias diminishes with age as visual precision increases and attentional control improves (Barutchu et al., 2009), leading to a gradual transition toward visual dominance in spatial tasks (Wille & Ebersbach, 2016). The ventriloquist effect—the mislocalisation of sound toward a spatially congruent visual source—illustrates how visual information dominates spatial judgements when modalities conflict (Alais & Burr, 2004). In children, however, the after effect of the illusion- a proposed example of sensory recalibration- is weaker and less consistent than in adults until around age 8 (Rohlf et al., 2020), suggesting that early multisensory spatial processing is governed by less stable weighting of sensory reliability.

2.2.3 Developmental Integration of Spatial and Temporal cues.

Integration across space and time continues to refine through middle childhood and adolescence. For spatial processing, studies show that visual–haptic integration reaches optimal (adult-like) weighting between ages 8 and 10 (Gori et al., 2008), while audiovisual spatial integration matures later, typically around age 12 (Nardini et al., 2016). Temporal integration, conversely, follows a more protracted course: even by early adolescence, the TBW remains broader than in adults (Hillock-Dunn & Wallace, 2012; Wallace & Stevenson, 2014). As previously mentioned, modality-specific trajectories suggest that unisensory maturation precedes multisensory optimisation. Burr and Gori (2012) propose that multisensory integration emerges only once the contributing senses provide sufficiently

reliable estimates to support statistically optimal fusion. Prior to this stage, children tend to rely on a single dominant sense, using others primarily for calibration. Bayesian causal inference models further explain developmental improvements in multisensory precision as the result of learning to evaluate cue reliability and causal structure (Körding et al., 2007; Ursino et al., 2014). Young children integrate cues indiscriminately, whereas older children selectively integrate those likely to share a common cause and suppress irrelevant or conflicting cues (Petrini et al., 2015; Verhaar et al., 2022).

In sum, the developmental trajectory of multisensory integration is best characterised by a gradual refinement in temporal alignment, spatial precision, and modality weighting. Infants begin with broad, redundant sensitivity, allowing them to detect amodal relations across senses. Through experience and neural maturation, these relations are progressively constrained: temporal binding windows narrow, and calibration processes align perceptual reference frames. Crucially, evidence from sensory deprivation underscores that these refinements depend on early bimodal experience. Without it, cortical areas repurpose to support unisensory compensation, and typical calibration hierarchies may never fully develop (Röder & Neville, 2003; Wallace & Stein, 2007).

2.3 Cross Sensory Calibration

During development, the senses do not initially combine optimally but instead undergo a period of mutual calibration. According to the cross-sensory calibration hypothesis (Gori et al., 2008; Gori, 2015), one sensory modality acts as a reference to fine-tune another, depending on which sense provides more accurate information for a given dimension. The origin of this theory is founded in both the philosophical propositions of Bishop George Berkely and the idea of modality appropriateness (Berkeley, 1983; Spence, 2016); vision calibrates haptic orientation perception; haptics calibrates visual size perception; and audition calibrates temporal perception (Gori, Giuliana et al., 2012). This process may explain why young children sometimes appear not to integrate multisensory

cues even when they possess both inputs: their perceptual systems are using one modality to correct systematic biases in another rather than integrating them. Empirical evidence supports this interpretation with Gori et al. (2008) finding that 5-year-olds rely predominantly on haptic information when judging size, whereas by age 8–10 they shift toward adult-like integration where visual and haptic cues are optimally combined. In a related study on orientation perception, younger children's performance was dominated by visual information, with integration emerging only once both senses reached sufficient unisensory precision (Gori et al., 2010). Therefore, cross-sensory calibration may be not a failure of integration but perhaps a developmental prerequisite for it. The brain must first establish accurate mappings between modalities before it can weight their inputs according to reliability, as predicted by Bayesian models of multisensory integration (Ernst & Bülthoff, 2004).

Research with populations who lack typical sensory experience provides critical insights into how calibration and integration depend on developmental experience. Studies of individuals who are congenitally blind or deaf demonstrate that the absence of cross-modal input leads to reorganised but functionally compensatory cortical networks (Röder et al., 1999; Röder & Neville, 2003). For instance, Röder et al. found that in congenitally blind individuals, visual cortical areas become recruited for auditory and tactile processing, enhancing spatial acuity and temporal discrimination (Röder et al., 2004; Gougoux et al., 2005). These cross-modal activations reflect functional plasticity but also highlight that the calibration processes typical in sighted individuals—such as vision calibrating spatial representations—cannot occur without early visual input. Similarly, congenitally deaf individuals often exhibit enhanced peripheral visual attention and motion sensitivity but show atypical integration between auditory and visual cues when hearing is restored (Bavelier et al., 2006). Together, these findings reveal that cross-sensory calibration is

experience-dependent, and that early sensory deprivation disrupts the typical hierarchy of modality appropriateness. Röder's work provides strong evidence that calibration and integration rely on early bimodal experience to establish intermodal correspondences. When this experience is absent, the sensory systems reorganise, but the developmental window for achieving typical integration appears limited.

2.4 Theories of Multisensory Integration across Development

Understanding how the ability to integrate sensory inputs emerges and matures requires synthesising evidence from behavioural, computational, and neurophysiological research. Theoretical accounts of multisensory integration during development have evolved from early dichotomies between *innate* and *experience-dependent* processes to more dynamic, probabilistic frameworks in which neural tuning, reliability estimation, and causal inference co-develop with sensory experience.

2.4.1 Early and Late Integration Accounts.

Historically, two broad theoretical positions have dominated debates on the ontogeny of multisensory integration: the early integration and late integration accounts. According to the early integration account, the infant nervous system is inherently multisensory, capable from birth of detecting redundancies across modalities such as synchrony, rhythm, and intensity (Bahrick & Lickliter, 2000; Lewkowicz, 2000). This view is supported by evidence that infants can recognise amodal relations—such as temporal synchrony in audiovisual speech (Dodd, 1979) or shape equivalence across vision and touch (Meltzoff & Borton, 1979)—well before the maturation of higher cognitive functions. The Intersensory Redundancy Hypothesis, proposed by Bahrick and Lickliter (2000), formalises this idea by proposing that redundant information presented across multiple senses enhances infants' attention to *amodal* features (e.g., tempo, duration) while suppressing attention to modality-specific details (e.g., colour, timbre). In contrast, the late integration account argues that the senses initially operate independently and must learn

through experience how to align information across modalities (Sloutsky & Napolitano, 2003; Gori et al., 2008). Early behaviours that appear integrative may instead reflect crossmodal matching or redundancy detection, not true integration. Under this view, the ability to combine cues in a statistically optimal fashion—weighting them by reliability—emerges only after the unisensory systems themselves are mature (Gori, Sandini et al., 2012; Burr & Gori, 2012). Empirical support for this account comes from findings that young children (under age 8) often rely on one dominant modality rather than combining cues, as seen in size and orientation tasks (Gori et al., 2008) and audiovisual localisation (Nardini et al., 2016).

More recent perspectives suggest these two accounts are not mutually exclusive. Rather, early redundant sensitivity provides a scaffold for learning causal associations, while later integration reflects the fine-tuned weighting of sensory estimates based on their reliability and task relevance (Ernst & Banks, 2002; Hillock-Dunn & Wallace, 2012).

2.4.2 The Intersensory Redundancy Hypothesis and Developmental Reweighting.

The Intersensory Redundancy Hypothesis (IRH) provides a unifying developmental framework that bridges early and late integration views. It proposes that when information is redundant across senses, infants preferentially attend to amodal properties (e.g., rhythm, synchrony), thereby promoting early learning of crossmodal correspondences. As cognitive and attentional systems mature, children gradually shift attention toward *modality-specific* features (Bahrick, Lickliter, & Flom, 2004). This attentional reweighting parallels the perceptual narrowing process described above, in which initially broad multisensory sensitivity becomes tuned to the contingencies that are most predictive within the child's environment (Lewkowicz & Ghazanfar, 2009). Thus, the IRH provides an early attentional mechanism that facilitates calibration and sets the stage for later precision-driven integration. Empirically, these ideas are supported by findings that infants' perception of

temporal and spatial synchrony enhances learning of audiovisual correspondences (Bahrick & Lickliter, 2000), while asynchronous or incongruent cues disrupt it (Lewkowicz, 1996). These effects are strongest for redundant rather than arbitrary cue pairings, suggesting that early multisensory perception is tuned by naturalistic correlations rather than learned symbolic associations.

2.4.3 Bayesian Causal Inference Models.

Modern theoretical accounts conceptualise multisensory integration as a process of probabilistic inference, where the brain estimates the most likely external cause of sensory events. Within Bayesian models (Ernst & Bülthoff, 2004; Ma et al., 2006), cues from different modalities are combined according to their *reliability*—that is, the inverse of their variance or uncertainty. The integrated percept represents a weighted average of these cues, yielding enhanced precision relative to any single modality. Developmental research applying these models reveals that while children are capable of *some* form of cue combination, they often fail to weight cues optimally. Gori and colleagues (2008; 2012) demonstrated that younger children over-rely on less precise modalities (e.g., touch for size perception) rather than combining cues based on reliability. Only by around age 8–10 do children achieve adult-like optimal integration. Similarly, Verhaar et al. (2022) found that children can infer whether two sensory signals arise from a common cause, but often fail to suppress irrelevant cues, indicating incomplete causal inference mechanisms.

These findings align with computational models showing that Bayesian learning of cue reliability requires both sensory experience and feedback. For instance, Ursino et al. (2014) modelled audiovisual integration in artificial neural networks and found that optimal Bayesian weighting emerged only after the system learned the statistical structure of the sensory environment. In human children, feedback on unimodal trials enhances subsequent integration performance, suggesting an experiential learning process similar to that observed in neural simulations (Negen et al., 2019). Thus, Bayesian frameworks account

for both the emergence and refinement of integration: early in development, broad crossmodal sensitivities support associative learning, while later, statistical learning mechanisms fine-tune cue weighting and causal inference.

2.5 Neurophysiological and Computational Accounts

At the neurophysiological level, multisensory integration develops through activity-dependent specialisation of cortical and subcortical circuits. Animal studies by Wallace and Stein (1997; 2007) demonstrated that neurons in the superior colliculus initially respond to single modalities but acquire multisensory responsiveness only after sufficient correlated cross-modal experience. Depriving animals of normal audiovisual input prevents this tuning, confirming that multisensory experience is necessary for establishing integration. Neuroimaging work in humans supports similar principles. Functional MRI and ERP studies show that multisensory integration engages a distributed network encompassing the superior temporal sulcus (STS), intraparietal sulcus, and superior colliculus (Calvert et al., 2004; Murray, Thelen et al., 2016). These areas exhibit developmental increases in both connectivity and response selectivity, paralleling behavioural improvements in multisensory tasks (Brandwein et al., 2011). Computationally, neural population models (Ma et al., 2006; Alvarado et al., 2008) describe how sensory neurons encode likelihood distributions and combine them through weighted summation, implementing a neural equivalent of Bayesian inference. Developmentally, these computations become more precise as synaptic efficiency, myelination, and cortical connectivity improve (Paus et al., 2001). Importantly, these models provide a mechanistic link between experience-dependent neural tuning and behavioural measures of multisensory reliability.

2.6 Integrative Developmental Theories

Synthesising behavioural, computational, and neurophysiological perspectives, contemporary developmental theories converge on a multistage model of multisensory development; first emerges early crossmodal sensitivity – infants detect broad

correspondences across modalities based on amodal features (Lewkowicz, 2000; Bahrick & Lickliter, 2000). Second, the capacity for cross-sensory calibration – during early and middle childhood, dominant senses calibrate weaker ones to align perceptual scales (Gori, 2015; Röder & Neville, 2003). Thirdly selective integration – as unisensory precision improves, children begin to integrate cues weighted by reliability (Ernst & Banks, 2002). Finally, optimal and flexible integration emerge – in adolescence and adulthood, integration becomes context-sensitive, guided by causal inference and task demands (Wallace et al., 2020; Verhaar et al., 2021).

This trajectory highlights that multisensory integration is not a singular ability but an emergent property of coordinated neural, perceptual, and cognitive systems. The developmental shift from calibration to integration marks a transition from establishing intermodal mappings to leveraging them for perceptual efficiency. Theoretical and empirical evidence collectively supports a dynamic, hierarchical model of multisensory development. Early crossmodal matching and redundancy detection provide a foundation for experience-dependent calibration, which, through learning and neural maturation, gives rise to true integration. Bayesian and causal inference frameworks offer quantitative descriptions of how the brain transitions from coarse correspondences to statistically optimal fusion. Neurophysiological and sensory deprivation studies underscore the importance of early multisensory experience in shaping integrative circuitry. In the absence of such experience—as in congenital blindness or deafness—cross-sensory calibration mechanisms reorganise, leading to atypical but adaptive perceptual networks (Röder & Neville, 2003). Together, these findings indicate that the development of multisensory integration emerges from the interaction between innate multisensory sensitivity, experience-dependent calibration, and the gradual refinement of neural systems that encode the temporal and spatial contingencies of the multisensory world.

3.0 Theories of the Development of Categories

The categorisation process is founded in the perception of shared distinctive features belonging to object categories such as their taxonomic classification- a hierarchical relationship between concepts such as terrier-dog-animal. This structure was pioneered by Rosch (1999) who proposed three levels of categorisation which varied in specificity. The highest level of classification is the superordinate level and is a general category such as 'animal', with the subsequent basic level class increasing in specificity and involves both high between-category differences, as well as high within-category similarities, such as a 'dog'. The lowest level is termed the subordinate level and involves specific subcategories such as a 'terrier'. Developmental research provides a critical lens for understanding how these theoretical mechanisms may emerge and transform across infancy and childhood. First this review will examine four prominent models of object categorisation—the Feature-Based Concept Theory, Prototype Theory, Exemplar Theory, and the Varying Abstraction Model (VAM)—with a specific focus on their basic principles, empirical support from developmental research, and their current standing in the field of developmental cognitive psychology.

3.1 The Feature-Based Concept Theory: the classical account

The Feature-Based Concept (FBC) approach, or classical view, proposes that category membership is defined by a set of *necessary and sufficient* features (Fodor, 1998; Medin & Schaffer, 1978). An object belongs to a category only if it possesses all defining properties—every member is equally representative. For example, the platypus would fail to qualify as a member of the category 'mammal' if mammals are strictly defined as animals that do not lay eggs. Developmentally, FBC predicts that infants first form categories based on salient perceptual features such as colour (Bornstein et al., 1976), orientation (Bomba, 1984), or physical parts (Quinn, 1999; Mandler, 2000; Arterberry & Bornstein, 2002). Indeed, even early in infancy, categorisation can be driven by such discriminable perceptual

information. Arterberry and Bornstein (2002) demonstrated that infants aged 6–9 months group static and dynamic objects according to shared features, indicating feature-based categorisation well before language acquisition. Similarly, Mandler and Bauer (1988) found that 12- to 20-month-olds first categorise at the *basic level* (e.g., "dog") rather than the more abstract *superordinate level* ("animal"), consistent with reliance on perceptual commonalities. However, the FBC model struggles to explain two pervasive phenomena: *typicality effects* (some category members are judged "better examples" than others) and *category boundaries* (Rosch & Mervis, 1975; Medin & Coley, 1998). Consequently, cognitive science has largely shifted toward more flexible, similarity-based accounts of categorisation.

3.2 Prototype Theory

The Prototype Theory (Rosch & Mervis, 1975; Rosch et al., 1976) proposes that categories are represented by a single abstracted *prototype*—the central tendency of previously encountered exemplars. New objects are classified according to their similarity to this prototype. This model accounts for graded category structure and typicality effects (i.e., as exemplars that are more similar to the prototype will be considered as more representative of the category; Minda & Smith, 2001). Developmental research indicates that infants possess the mechanisms necessary for prototype formation. Using habituation–dishabituation paradigms, Bomba & Siqueland (1983) and Younger & Gotlieb (1988) showed that infants as young as 7–10 months abstract a prototype from variable exemplars and prefer it even when never directly seen. Similarly, Quinn et al. (2001) demonstrated that 3-month-olds categorise 2D shapes by global contour rather than local details, suggesting abstraction of central tendencies in early perception. Quinn and Johnson (2000) further observed that 2-month-olds could categorise pictures of mammals versus non-mammals, implying sensitivity to superordinate structure. Further empirical evidence suggests categorisation is refined by exposure and experience. Needham et al. (2005)

reported that the variability of exemplars profoundly affects infants' category learning: when exemplars are highly variable, infants generalise more inclusively, consistent with averaging across instances. Bornstein and Arterberry (2010) extended this showing that 12- to 30-month-olds progressively form more inclusive categories (e.g., "animals") before narrower ones ("dogs"), reinforcing Rosch's (1999) hierarchy of *superordinate–basic–subordinate* levels. Prototype formation has also been linked to conceptual abstraction: from around 4 years of age, children recognise "family resemblance" between category members, relying on shared central features rather than rigid definitions (Landau et al., 1998b; Lopez et al., 1992). Overall, prototype-based processes appear early and provide a mechanism for efficient generalisation.

3.3 Exemplar Theory

In contrast to the prototype theory, Exemplar Theory (Medin & Schaffer, 1978; Nosofsky & Palmeri, 1997) posits that categories are represented not by an abstracted mean but by stored *individual instances*. Categorisation occurs through comparison of a novel object to all remembered exemplars; membership is determined by summed similarity. Evidence from infancy supports exemplar-sensitive processing. Oakes and Spalding (1997) found that infants' category inclusiveness depends on exemplar variability: exposure to a diverse set of exemplars yields broader generalisation, suggesting retention of multiple specific instances. Mareschal, French, and Quinn (2000) further demonstrated that manipulating exemplar similarity changes category boundary exclusivity—categories with greater within-set variability (e.g., diverse dogs) produce more inclusive generalisations than less variable ones (e.g., cats). These findings imply that infants store and compare distinct exemplars rather than relying solely on averaged prototypes. With development, the balance between prototype abstraction and exemplar retention appears susceptible to change. Hayes & Taplin (1993) showed that younger children rely more heavily on prototype information, whereas older children and adults flexibly combine both sources.

Thus, exemplar theory complements prototype theory by explaining variability, contextual sensitivity, and instance-specific learning observed throughout development.

3.4 The Varying Abstraction Model

The long-standing prototype–exemplar debate has given rise to integrative frameworks such as the Varying Abstraction Model (VAM; Vanpaemel & Storms, 2008). VAM conceptualises category representations along a continuum of abstraction—from fully concrete exemplars to fully abstract prototypes, thereby bridging exemplar and prototype models. Category learning can involve multiple sub-prototypes or clusters of exemplars, allowing graded abstraction tuned to task demands. Originally developed to model adult categorisation, VAM's flexibility offers strong explanatory potential for developmental change. As memory capacity, attentional control, and experience expand, children may transition from reliance on specific exemplars toward more abstract, efficient representations. Bayesian model comparisons further support the plausibility of partial abstraction (Lee & Vanpaemel, 2008; Vanpaemel & Storms, 2010). In this view, developmental changes may reflect a continuous adjustment in abstraction level rather than a categorical shift between mechanisms.

3.5 The Developmental Trajectory of Conceptual Versus Perceptual Categorisation

A major theme emerging from developmental research is the transition from perceptual to conceptual categorisation. Early in life, categorisation is largely perceptually driven infants rely on surface similarities such as contour and colour (Quinn et al., 2001). As cognitive and linguistic systems mature, children incorporate conceptual and functional knowledge, such as causal relations or intended use (Mandler, 2008; Oakes & Rakison, 2003). Thus, perceptual, and conceptual processes are interdependent, with language labels and semantic knowledge sharpening category boundaries (Althaus & Mareschal, 2014). Neuroscientific accounts further suggest that categorisation arises through dynamic interactions between bottom-up sensory inputs and top-down predictive feedback. Bottom-

up pathways convey perceptual evidence, while top-down signals from higher cortical areas integrate prior experience to resolve ambiguity (Riesenhuber et al., 2009; Rauss & Pourtois, 2013). Developmental neuroimaging indicates increasing cortical specialisation and connectivity supporting abstraction and generalisation, with adults showing category specific activation in temporal and prefrontal regions (Schendan & Ganis, 2015). These findings align with hierarchical predictive coding frameworks, suggesting that the maturation of feedback circuits may underlie the shift from perceptual grouping to conceptual representation.

3.6 The Development of Category Learning in Multisensory Contexts

Traditional theories of categorisation were largely developed using unimodal stimuli, typically visual objects presented in isolation. However, real-world learning rarely occurs within a single sensory modality. From infancy, objects are experienced through coordinated visual, auditory, and tactile input, suggesting that category formation is inherently multisensory (Gibson, 1969; Murray et al., 2016; Lewkowicz, 2014). Developmental research increasingly demonstrates that multisensory information plays a fundamental role in guiding attention (Bahrick & Lickliter, 2000; Talsma et al., 2010), highlighting relevant features (Deng & Sloutsky, 2016; Kruschke, 2001), and stabilising category representations (Murray & Wallace, 2011; Newell et al., 2023). These findings challenge purely similarity-based, unimodal accounts and instead indicate the necessity of frameworks which incorporate crossmodal interactions across development

3.6.2 Early Mechanisms of Category Learning.

Infants form perceptual categories within the first months of life (Quinn & Eimas, 1996; Quinn et al., 2001). Within this early period, category learning is driven primarily by bottom-up sensitivity to perceptual regularities. Infants extract structure from recurring features and co-occurrence patterns, forming flexible but relatively shallow representations that support recognition and limited generalisation (Quinn & Eimas, 1994; Deen et al.,

2017; Kosakowski et al., 2022). Statistical learning accounts (Saffran et al., 1996) provide a complementary mechanism through which early category learning may emerge, proposing that infants track regularities not only within but also across sensory modalities. This capacity for cross-modal statistical learning enables the gradual abstraction of category structure from richly multisensory environments (Bahrick & Lickliter, 2000; Murray & Shams, 2023). Crucially, this framework implies that early category representations are not strictly modality-specific, but are shaped by structured, often multisensory, input, providing a developmental basis for the integration of information across sensory systems (Johnson, 2011; Johnson, 2001). One influential developmental account of these effects is the Intersensory Redundancy Hypothesis (IRH; Bahrick & Lickliter, 2000; 2004), which proposes that temporally synchronous, redundant information across modalities selectively directs attention to amodal properties (e.g., rhythm, tempo), thereby facilitating early learning. In this way, multisensory redundancy is thought to scaffold category formation before the maturation of selective attention and executive control.

3.6.3 Neural Foundations of Category Learning.

Neuroimaging evidence further supports the early emergence of category-relevant structure in the brain. By 2–9 months of age, the cortex already exhibits category-selective responses for faces, scenes, and bodies in regions homologous to adult areas such as the Fusiform Face Area (FFA), Parahippocampal Place Area (PPA), and Extrastriate Body Area (EBA), although these representations are more broadly tuned and continue to refine with experience (Deen et al., 2017; Kosakowski et al., 2022; 2024). These findings are consistent with interactive specialization accounts (Johnson, 2011; Gauthier & Nelson, 2001), which propose that coarse domain-relevant biases are present early in development and become increasingly specialised through experience and network reorganisation. Within this framework, early category representations are not strictly modality-specific but

are shaped by structured, often multisensory, input, providing a developmental basis for the integration of information across sensory systems (Johnson, 2011).

3.6.4 The Role of Attention and Language in the Development of Category Learning.

As development progresses into early childhood (2–5 years), children's category learning is increasingly shaped by interactions between attentional learning and linguistic labels. Experience naming objects tunes attention toward diagnostic shape with children generalising novel count nouns by shape more than texture or size, and experimentally training this bias accelerates later vocabulary growth (Landau et al., 1988c; Smith et al., 2002; Perry et al., 2011). These findings link object labels, attention, and category formation as inherently related phenomena; from a multisensory perspective, labels may function to associate information across modalities, thereby stabilising category representations and supporting generalisation across contexts (Deng & Sloutsky, 2016; Matusz et al., 2017). In this way, language may not merely reflect category structure but actively contributes to their development by guiding attention and reinforcing cross-modal associations.

3.6.5 Multisensory Facilitation of Category Learning and Potential Constraints.

A growing body of research demonstrates that multisensory input can enhance category learning, particularly when information across modalities is congruent (Heikkila & Tiippana, 2016; Lewkowicz, 2014). In adults, training with congruent audiovisual stimuli improves subsequent visual categorisation more than unimodal training, whereas incongruent pairings provide little benefit (Kim et al., 2008). More broadly, multisensory training has been proposed to more closely approximate natural learning environments, enhancing neural salience and stabilising representations (Shams & Seitz, 2008).

Developmental evidence suggests that these benefits emerge earlier in indirect forms. Children as young as 3–4 years show improved learning and encoding when information is presented multisensorily (e.g., Nardini et al., 2010; Bahrick & Lickliter, 2014), and by early school age, audiovisual input enhances category learning relative to unimodal input (Broadbent et al., 2017). However, direct evidence for multisensory training leading to improved subsequent categorisation in 2–5-year-old children remains limited. Instead, multisensory input appears to support foundational processes such as attention, binding, and mapping (Gogate & Bahrick, 2001; Gogate et al., 2006; Matatyaho & Gogate, 2008), which may scaffold later category learning.

Importantly, multisensory facilitation is not uniform. Empirical evidence indicates that incongruent or weakly aligned (temporally or spatially) cues can impair learning, and that the benefits of multisensory input depend on task demands, developmental stage, and available attentional resources (Lewkowicz, 2014; Matusz et al., 2017). While the IRH provides a compelling account of early attentional biases toward redundant, amodal information, it is less explicit in explaining how learners come to process modality-specific features or integrate non-redundant cues, both of which are essential for mature category representations (Smith & Medin, 1981; Ashby & Maddox, 2005; Sloutsky, 2010). A comprehensive account of multisensory category learning must therefore accommodate both facilitative and interfering effects of cross-modal input.

3.6.6 Developmental Trajectory of Category Learning.

Beyond early childhood, continued improvements in executive control and metacognition support increasingly flexible category learning. In infants, categories emerge from perceptual clustering of recurring features (e.g., bounding contours, spatial relations) and are detectable in looking preferences (Quinn & Eimas, 1994) and early neural selectivity (Denn et al., 2017; Kosakowski et al., 2022). These categories are flexible but shallow: they support recognition and rudimentary generalization but are highly dependent

on salient features and context. In the first years of life with rapid growth in vocabulary, object naming trains attention to task-relevant dimensions (e.g., shape for solid object kinds), fostering more stable categories and faster novel category acquisition (Smith, 2009). Labels function both as additional cues and as organizers of similarity, with their role shifting as semantic networks expand. In school age children, research suggests the improvements in executive control and metacognition increase success on rule-based tasks (Reetzke et al., 2016); children show adult-like gains when categories are dimensionally separable and feedback is immediate (Maddox et al., 2003; 2005). On probabilistic or information-integration tasks, performance continues to improve into adolescence as procedural learning and integration of multiple cues become more reliable (Reetzke et al., 2016; Roark et al., 2023). In adults, delayed feedback selectively harms information-integration category learning (procedural/striatal) while sparing rule-based learning; model fits show learners switch to (inefficient) hypothesis testing under conditions of delayed feedback (Maddox et al., 2003, 2005). Developmentally, children show parallel patterns with performance on rule-based, dimensionally separable tasks improving with age, whereas probabilistic and information-integration learning shows a more protracted trajectory (Reetzke et al., 2016; Roark et al., 2023).

Across development, object category learning mechanisms evolve from perceptual clustering in infancy to flexible, multi-format representations in adulthood. Infants rely on bottom-up statistical regularities and salient perceptual features to form coarse categories (Arterberry & Bornstein, 2002; Bahrick & Lickliter, 2004). During early childhood, attentional tuning and labels support more efficient category learning, with a gradual shift from solely perceptual to interdependent perceptual and conceptual bases (Bornstein & Arterberry, 2010). Middle childhood brings improvements in feedback utilization and rule-based reasoning, enabling more abstract and flexible categorisation strategies (Rabi &

Minda, 2014; Hayes & Taplin, 1993). By adulthood, learners can flexibly combine exemplar storage (Lee & Vanpaemel, 2008), prototype abstraction (Vanpaemel & Storms, 2008), rule-based reasoning (Maddox & Ashby, 2004), and multisensory information (Lacey & Sathian, 2014; Lewkowicz, 2014; Matusz et al., 2017; Murray, Lewkowicz et al., 2016), supported by mature attentional (Deng & Sloutsky, 2016), linguistic (Smith, 2009), and neural systems (Jao et al., 2015; Murray, Thelen et al., 2016). These developmental trajectories provide essential context for understanding how multisensory object categorisation emerges and stabilizes, and they form the theoretical foundation for the empirical work presented in this thesis.

3.7 Current Standing

Research in infants and children demonstrates that the ability to form object categories emerges early and undergoes progressive refinement. Developmental research has moved beyond the constraints of the classical Feature-Based Concept Theory, finding support for the flexibility of similarity-based models. The current standing in the developmental cognitive field is not one of a winner-take-all contest between Prototype and Exemplar theories, but rather a recognition that both abstraction and instance-specific memory are active from infancy. Across development, increasing cognitive resources and experience promote more flexible, context-sensitive categorisation, consistent with continuum models such as VAM. Moreover, the integration of perceptual and conceptual processes, supported by developing neural circuitry, underscores categorisation as a hierarchical, adaptive capacity central to cognitive development. Furthermore, these highlight a critical limitation in existing accounts of category learning in children, while the roles of perceptual similarity, attention, and language are well established, the contribution of multisensory processes remains comparatively under-specified. Multisensory input cannot be assumed to be uniformly beneficial; rather, its effects may depend on congruency, task demands, and developmental stage.

4.0 Multisensory Object Perception across Development: Recognition and Categorisation

The perception of objects in everyday life rarely occurs through a single sensory domain. Rather, humans continuously combine and compare information across vision, audition, and haptics to encode, recognise and categorise objects. Over development, these abilities undergo profound change, shifting from early sensitivity to redundant, amodal cues in infancy to more sophisticated, reliability-weighted integration strategies in later childhood. By around four years of age, children already benefit from redundant audiovisual cues in spatial decision-making tasks, showing faster and less variable localisation than when either modality is presented alone (Negen & Nardini, 2015; Nardini et al., 2016). However, the combination of auditory and visual information in young children is often suboptimal compared to adults, suggesting that although children can pool cues, their integration strategies are not yet fully efficient. Converging evidence suggests that early in development, sensory systems may interact through processes of calibration, whereby one modality (often vision) supervises or aligns another, before adult-like integration emerges (Gori et al., 2008; Burr & Gori, 2012).

Adult object perception and categorisation are fundamentally multisensory: people routinely combine information from vision, audition, and haptics to recognise and classify objects efficiently and flexibly. In adulthood, crossmodal processing and multisensory integration interact with object recognition and categorisation being supported by shared, modality-independent representations and flexible cue combination strategies that depend on task demands and sensory reliability (Woods & Newell, 2004; Lacey & Sathian, 2014). This review will now consider empirical evidence for crossmodal object recognition, multisensory object recognition, crossmodal object categorisation, and multisensory object categorisation in adults, infants and children will be conducted to assess the current understanding of how multisensory object perception emerges.

4.1 Crossmodal Object Recognition

Adults display robust crossmodal transfer of object identity and shape information between vision and haptics. Behavioural studies show that both modalities are sensitive to viewpoint when recognising unfamiliar three-dimensional objects, but that crossmodal recognition can be viewpoint-independent, implying shared representational structures (Newell et al., 2001; Lacey et al., 2007). Crossmodal matching is also sensitive to changes in size: both unimodal and crossmodal recognition performance declines similarly with size transformations, consistent with a common, size-sensitive shape code underlying both senses (Craddock & Lawson, 2009). Neuroimaging confirms that tactile object perception activates ventral visual stream areas, especially the lateral occipital complex (LOC), in a modality-independent fashion, indicating that haptic information is recoded into a visual-object format for recognition (Amedi et al., 2001; Pietrini et al., 2004). Thus, in adults, crossmodal recognition is supported by abstract, modality-independent object representations, particularly for familiar objects, and by efficient transformations that align sensory input to these representations.

The earliest evidence of crossmodal object processing comes from neonates who demonstrate rudimentary transfer of shape and texture information between touch and vision. Neonates who explore an object haptically look longer at the matching visual shape compared to a novel one, indicating cross modal transfer of information for the purpose of recognition (Meltzoff & Borton, 1979; Streri & Gentaz, 2004; Streri & Gentaz, 2004; Sann & Streri, 2007). However, this ability is asymmetric: haptic-to-visual transfer emerges at birth, whereas visual-to-haptic transfer develops more slowly. This asymmetry likely reflects early haptic dominance in shape processing, differences in exploratory competence, and sensory maturation.

In preschool and early school-aged children, crossmodal object recognition shows strong effects of object familiarity and exploration strategies. Using familiar everyday

objects, Purpura et al. (2018) found that children recognised objects most accurately when exposure to both senses occurred sequentially, but that accuracy substantially improved with age, particularly when comparing performance of children aged 4–5 to those of 8+ years. For unfamiliar objects, Kalagher and Jones (2011b) showed that 2.5–4-year-olds performed at chance when required to recognise objects visually after haptic exposure, whereas 5-year-olds performed reliably above chance. Their subsequent analyses revealed that younger children's haptic exploration was unsystematic, preventing them from extracting shape-defining information. By around age five, efficient exploration strategies support successful crossmodal transfer, marking a key developmental step toward the adult pattern of robust, viewpoint-tolerant crossmodal recognition.

4.2 Multisensory Object Recognition

When object information is presented simultaneously across senses, adults typically integrate cues in a near-optimal manner, combining them in proportion to their reliabilities to produce a more precise perceptual estimate than either modality alone (Ernst & Banks, 2002). This has been demonstrated in visuo-haptic shape judgments (Helbig & Ernst, 2007) and audiovisual spatial localisation (Alais & Burr, 2004). Crucially, integration depends on causal inference: cues are fused only if they are likely to originate from the same object, with spatial, temporal, and semantic congruency acting as key determinants (Körding et al., 2007; Helbig & Ernst, 2007). Beyond low-level cue combination, semantic associations powerfully shape object recognition. Characteristic sounds facilitate visual search and object identification, even under degraded visual conditions (Laurienti et al., 2004; Chen & Spence, 2010; Iordanescu et al., 2008; 2010). Functional imaging and TMS studies implicate the LOC as a convergence site for visuo-haptic shape information, with crossmodal enhancement depending on familiarity and experience (Jao et al., 2015; Amedi et al., 2007).

Evidence in children for simultaneous cue integration emerges somewhat later than early crossmodal transfer. By around age four, children exhibit the behavioural benefits of audiovisual integration in spatial localisation tasks (Nardini et al., 2016), indicating an ability to exploit redundant cues for improved performance. However, children's integration is suboptimal compared to adults: they often fail to weight cues by their relative reliabilities and may over-weight synchronous but uninformative signals (Petrini et al., 2015). Electrophysiological studies show that multisensory ERP signatures, such as mid-parietal amplitude enhancements, are reduced in young children compared to adults, reflecting immature integration at neural levels (Brandwein et al., 2011).

Visuo-haptic integration follows a protracted trajectory. Young children rely predominantly on a single dominant modality, often haptics for size and vision for orientation, rather than combining cues optimally (Gori et al., 2008; Gori et al., 2010). The calibration hypothesis proposes that during early childhood, one modality (typically vision) serves as a reference for calibrating another (typically touch), delaying true integration until approximately 8–10 years of age. Audiovisual semantic congruency also shows developmental change: while infants and young children are sensitive to temporal synchrony and simple correspondences, the top-down modulation of recognition by semantic congruency—such as faster identification of objects paired with characteristic sounds—emerges more strongly in later childhood and adolescence (Heikkilä & Tiippana, 2016). Audio-haptic integration, in the absence of vision, is particularly late developing. Studies show that optimal integration of auditory and haptic cues for size discrimination does not emerge until early adolescence (Petrini et al., 2014; Scheller et al., 2019, 2021). Younger children often fail to integrate these cues, relying instead on a single modality, which underscores the dominant role of vision as an integrative anchor during early development.

4.3 Crossmodal Object Categorisation

Critically, adults can transfer category knowledge across modalities, upon learning to categorize novel objects through one sense, adults can categorize new exemplars through a different sense without explicit training, reflecting modality-independent category representations (Yildirim & Jacobs, 2013). This crossmodal transfer mirrors findings in recognition tasks and supports the view that mature multisensory systems form shared prototypes or category structures that transcend individual modalities. Crossmodal training can even make within-modality recognition more invariant—for example, visuo-haptic training can make subsequent visual recognition more robust to viewpoint changes (Lacey et al., 2010). Thus, adult object categorisation relies on flexible, shared representations and optimal cue combination, with strong semantic and experiential modulation.

Even before word knowledge is robust, infants use intersensory redundancy (e.g., temporal synchrony between object motion and speech) to bind labels to objects, scaffolding early category formation across senses (Gogate & Bahrick, 1998, 2001; Bahrick & Lickliter, 2004, 2006). Such crossmodal bindings, or crossmodal associations, support the transition from perceptual groupings to conceptual categories. When categories are defined by features available across senses (e.g., a visual pattern *and* an auditory timbre), school-age children can use the crossmodal regularities to learn categories, though younger groups benefit most from redundant bimodal information (e.g., 5-year-olds show largest AV advantage) (Broadbent et al., 2018a). Related work comparing how children weigh visual vs. auditory features in category decisions also shows age-related shifts in reliance on modality and feature type (Berger & Donnadieu, 2008).

The ability to form and use categories across modalities builds on early perceptual capacities. By 3–4 months of age, infants can categorise visual stimuli based on features such as shape, colour, and spatial relations (Bornstein et al., 1976; Quinn, 1987, 1994). Through intersensory redundancy, infants learn to bind labels and perceptual events:

synchronous audiovisual information supports the formation of object–label associations, scaffolding later crossmodal category structures (Gogate & Bahrick, 1998, 2001; Bahrick & Lickliter, 2004, 2006). In older children, crossmodal categorisation depends strongly on feature salience and exploratory competence. When diagnostic information is equally available in vision and touch, children gradually become able to generalise category structures across modalities, though this ability lags behind recognition (Berger & Donnadieu, 2008). These findings suggest that while the foundations of crossmodal categorisation are laid in infancy, their full deployment depends on the maturation of attentional control and multisensory representations.

4.4 Multisensory Object Categorisation

In contrast to crossmodal transfer, multisensory categorisation tasks present information from multiple modalities simultaneously. Adults display efficient multisensory category learning, particularly when auditory and visual cues provide redundant or complementary diagnostic information. Multisensory training improves learning rates and retention compared to unimodal training (Shams & Seitz, 2008). Visuo-haptic similarity spaces for natural objects are highly correlated, suggesting that category boundaries are shared across modalities (Gaissert & Wallraven, 2012). Audiovisual categorisation benefits from semantic congruency, and adults can learn novel category rules spanning independent auditory and visual dimensions (Laurienti et al., 2004; Viganò et al., 2021). Multisensory experiences can induce plastic changes in sensory cortices, indicating that multisensory learning reshapes sensory representations themselves (Knöpfel et al., 2019).

Developmental patterns reflect both the benefits of redundancy and constraints imposed by immature integration. Using the Multisensory Attention Learning Task (MALT), Broadbent et al. (2018b) found that 6–10-year-old children acquired audiovisual categories faster than unimodal categories, with the greatest benefit observed in 5–6-year-olds. This indicates that redundant audiovisual cues can compensate for less efficient

unimodal processing in younger children. Older children were more flexible, performing well even under unimodal conditions, reflecting increasing unimodal competence and attentional control. In the visuo-haptic domain, Broadbent et al. (2020a) showed that younger children learn categories best from haptic-only or visuo-haptic training, depending on the relative salience of diagnostic features. For example, when compressibility (a haptic feature) and pattern frequency (a visual feature) are diagnostic of category membership, younger children relied on haptics, while older children integrated both cues and performed equivalently across modalities. These results illustrate that what appears as a "multisensory advantage" in children may sometimes reflect strategic reliance on the most salient or reliable cue rather than true integration.

Finally, semantic congruency begins to shape multisensory categorisation in late childhood. Heikkilä and Tiippana (2015) found that semantically congruent auditory or verbal cues paired with visual objects enhanced memory performance in 8–12-year-olds, paralleling adult effects. This indicates that top-down semantic influences on category learning emerge later than basic redundancy-based effects.

5.0 Outline and Scope of this Thesis

The human brain possesses remarkable capacity to organize sensory information from competing sources into meaningful categories, enabling efficient recognition and interaction with objects in daily life. This ability is especially impressive given the spatial and temporal complexity of our multisensory environment. The trajectory and nature of how multisensory information informs perceptual judgements in childhood is under some debate, yet in adulthood it forms the foundation of how we experience the world. Although recent research has greatly advanced our understanding of how multisensory and cross-modal experience contribute to object perception across multiple sensory modalities—highlighting early cross-modal interactions that facilitate efficient decision-making—our understanding of how these sensory representations integrate to form stable category

memories remains limited. In the past, investigations into object perception and categorisation have largely been conducted in a segmented fashion, with a focus on infancy and adulthood, and examining higher-order cognitive, attention, and unimodal low-level perceptual influences independently. In order to advance our existing understanding of multisensory categorisation in childhood as well as advance towards a comprehensive theory of the development of multisensory category learning, extending beyond unimodal frameworks, empirical research is needed to account for how category information is acquired across sensory modalities, how attentional mechanisms may prioritise cross-modal input, and how these processes evolve across development. In particular, further work is needed to specify how multisensory experiences influence not only the acquisition of novel category structures, but also their stability and ability to generalise category representations over time. Addressing these questions is central to the empirical work presented in this thesis.

To provide novel insights into the formation of multisensory object categories in childhood, a series of experiments have been designed to examine the relative contribution of auditory, visual, and haptic information to children's emerging capacity to acquire novel multisensory object categories. Chapter 2 explores how audiovisual and contextual information contribute to the categorization of familiar objects in children aged 5 to 13 years. Chapter 3 compares children aged 5-13 years and adults' capacity to acquire novel dynamic audiovisual object categories and examines the characteristics of the resulting memory representations. Subsequently, Chapter 4 compares visuohaptic category learning with unisensory conditions (visual-only and haptic-only) in children aged 4 to 13 years, evaluating the influence of age and learning modality on cross-modal categorization and generalization. In the final empirical chapter (Chapter 5), eye-tracking is employed during visual category learning to compare in children and adults the potential effects of selective

attention during acquisition of novel visuohaptic object categories when ones testing modality is un/known to be visual or haptic. The broader theoretical implications of these behavioural findings are discussed in Chapter 6, alongside proposed directions for future research.

Chapter 2

The Role of Audiovisual Object Features and Scene Context on Children's Ability to Categorise Familiar Objects

Chapter 2

Abstract

Children's ability to visually categorise objects emerges early in development. However, the use of relevant features from other modalities, or contextual scene cues, on categorisation across development is poorly understood. In particular, children's categorisation may benefit from any associated cue to categorisation or may be specific to the availability of object-relevant sensory information. Furthermore, a benefit of additional information on categorisation may depend on age. To investigate this, we tested children between the ages of 5 to 13 years on their categorisation of familiar objects, which were presented either as images-only (V), sound-only (A) or bimodal audio-visual (AV) objects. In addition, the objects were presented in isolation or within a scene that was either semantically congruent or incongruent to the object category. First, we found that categorisation accuracy and response time performance improved with age, as expected. Children were better able to categorise images than sounds of objects, but there was no overall advantage for combined AV information over images only on either accuracy or response times. A semantically congruent context did not further improve categorisation accuracy, although response times were fastest to objects presented within a congruent scene context, particularly when an object's sound was presented. Our findings suggest that visual information dominated the categorisation of familiar objects over sounds in children, with no evidence that these cues are combined at any age. However, children's categorisation of objects benefitted from associative cues such as background scene context, particularly when the object information was impoverished. Our findings have important implications for understanding how children adaptively use available sensory information in the categorisation of familiar objects.

Chapter 2

Introduction

The ability to recognise and categorise objects is a fundamental cognitive skill that emerges early in development (Quinn & Eimas, 1986). Infants as young as 3-4 months are already capable of categorising living and non-living objects (Mandler et al., 1991; Behl-Chadha, 1996; Peykarjou et al., 2023) and even within these broad taxonomies, young infants can further distinguish between similar objects such as cats and dogs (Quinn, Eimas & Rosenkrantz, 1993; Quinn, Eimas & Tarr, 2001), horses and zebras (Eimas & Quinn, 1994) and artefacts such as chairs (Quinn & Eimas, 1996). Moreover, findings from neuroimaging studies have supported the idea that the developing brain organises visual information into categories from an early age (O'Doherty et al., 2026; Kosakowski et al., 2022; Spriet et al., 2022; Xie et al. 2022) with different developmental trajectories across object categories (Yan et al. 2024), depending on regional maturation rates (Lebenberg et al., 2019; Ellis et al., 2021) and experience (Deen et al. 2017).

Several studies have provided insights into the nature of children's visual object representations across different ages. Studies of visual object perception suggest that young infants rely more on local object features, and the objects are recognised on the basis of their global shape from around the age of 24 months (Smith, 2009; Gershkoff & Smith, 2004; Augustine et al., 2011). More complex perceptual processes, such as invariant object recognition, is evident later in development, from the age of six years (Bova et al., 2007). Relatedly, a developmental shift occurs from perceptually based categorisation at around 4 months of age (Mareschal et al., 2005) to more thematically based object associations at around the age of two (Nelson, 1988; Blewitt, 1994; Bauer & Mandler 1989). In older children, category judgements are more informed by structured taxonomic representation (Bjorklund, 1985; 1987; Deak & Bauer, 1995; Horton & Markman, 1980). Substantial progress has therefore been made towards a better understanding of object categorisation in children, at least based on visual information (Ayzenberg & Behrmann, 2024).

Despite strong evidence that visual object categorisation improves with development, what is less clear, is how other relevant information supports object categorisation across childhood. For example, many objects make sounds, such as a dog's bark, and these sounds may be diagnostic for category membership (e.g. Brunel et al. 2013). Although categorisation in children has been mainly studied based on visual information (see e.g. Mareschal & Quinn, 2001), it is known that sound can affect categorisation from the early stages of development (Patterson & Werker, 2002; Miller, 1983). Furthermore, objects are rarely viewed in isolation, and their context may provide an associative cue to the object's category. Thus, an indoor, domestic scene may facilitate the recognition of a kettle but not of a water hydrant due to differences in their semantic congruency (Biederman et al., 1982; Palmer, 1975; Boyce et al., 1989; see also Hollingworth & Henderson, 1999).

In children, the adaptive use of relevant, multisensory features or contextual cues for object categorisation may be constrained by the development of perceptual (Gori et al., 2008; Nardini et al., 2008; Helo et al., 2014; Luna et al., 2008) and cognitive processes (Newcombe et al., 1977; Ofen et al., 2012) as well as experience or prior knowledge (Darby et al., 2021; Watson et al., 2025). Indeed, compared to perception in adults, evidence for a benefit of multisensory information on perception in children is not consistent (Matusz et al., 2015). On the one hand, children's ability to integrate redundant multisensory cues does not appear to be optimal until later stages of development, i.e. after age of 8 years (e.g. Gori et al., 2008). However, studies by Broadbent and colleagues suggest that category learning in young children is enhanced by combined audio-visual cues relative to either sensory modality alone (Broadbent, White et al., 2018; Broadbent et al., 2020). Moreover, Kirkham et al. (2019) reported a greater benefit of audio-visual relative to unisensory exposure on object category learning specifically in younger (i.e. 5-year-old) but not older (7- or 10-

year-old) children. The focus of these studies was, however, on how multisensory information is used in the formation of novel object categories, therefore the question of whether children can flexibly use cross-modal or context-based cues to enhance their categorisation of known objects remains open. In other words, our understanding of the multisensory nature of object category representations in semantic memory, and the implementation of these representations for object recognition across development, remains relatively poor.

Studies conducted with adult participants suggest a benefit of audio-visual congruent information on the recognition of familiar objects (e.g. Lehmann & Murray, 2005; Thelen et al., 2015; Chen & Spence, 2010; Li et al., 2020). Studies of familiar object perception in children, however, suggest more nuanced effects across development. For example, Thomas, Nardini and Mareschal (2017) reported that children's categorisation of the sounds of familiar objects is enhanced when a semantically congruent image of the object is also presented (i.e. a woof sound and image of a dog) relative to an incongruent (e.g. a lion) or control image. Interestingly, a cost in performance by the presence of an incongruent image was found in older children only (8-9 years) but not younger children (6-7 years), suggesting that semantically relevant object information is processed differently in middle childhood. Indeed, sensory dominance may play an important role in children's ability to perceive objects. For example, even when perceptual salience is equated across vision and audition, sounds can dominate perception in audiovisual tasks in children as young as four-years (Robinson & Sloutsky, 2004a; Sloutsky & Napolitano, 2003). In contrast, Broadbent, White et al., (2018) reported that category learning was worst to auditory-only stimuli in 6-year-olds, although learning performance benefitted from audiovisual stimulation in children aged between 6- and 10-years old. To account for these findings, the authors ruled out differences in auditory working memory across age groups,

or salience differences across modalities, and instead suggested that younger children may rely more on visual information when acquiring category knowledge. Despite evidence for sensory dominance in younger children, by 8 years of age, semantically congruent audiovisual information benefits object recognition relative to incongruent information (see also Heikkila & Tippana, 2016).

The role of contextual cues in object categorisation in children has not yet been firmly established although there is some suggestion that scene gist is perceived from an early age (Duh & Wang, 2014) and that semantic cues from a scene affects object perception by 4-years of age (Oehlschlaeger & Vo, 2020). Thus, the effect of scene context may be similar across all children although competition for attention from multiple sources of information may further challenge perception particularly in younger children (Lickliter & Bahrick, 2004). As such, a benefit for semantically relevant scenes on the visual or auditory categorisation of objects may improve with age.

The current study was designed to address several questions relating to children's ability to use object-based or semantic context cues for the purpose of categorising objects: a) do combined auditory and visual cues enhance the categorisation of familiar objects relative to unisensory cues in children; b) does a semantically congruent scene enhance the categorisation of objects relative to no context or an incongruent context in children and c) is there a further benefit on children's categorisation performance by combining all relevant (object-based or scene-based) cues? In addition, we also ask whether the effect of any of these object-based or scene-based cues on categorisation is influenced by development.

To that end, we tested children aged from 5 to 13 years on their ability to categorise familiar animals based on their image alone, sound alone or by combining the image with the sound. In addition, these objects were presented within different background contexts that were either semantically congruent (a lion in a jungle) or incongruent (a lion in a

farmyard) or no background scene was presented. This age range was selected in order to capture potential age-related differences in how children utilise both contextual (Thomas et al., 2017) and object information from multiple sensory modalities (Heikkila & Tippana, 2016). By examining children's performance across combinations of object modality and scene context, we aimed to clarify how bottom-up sensory information and top-down semantic congruency dynamically interact during a critical period of cognitive and perceptual development.

Method

Participants

We recruited 181 children (96 boys and 85 girls) to take part in this study from four primary schools (children from 1st to 6th class) within the greater Dublin area and surrounding counties in Ireland. An a priori power analysis determined that a sample size of 144 participants would be required to detect a small effect 0.25 with 90% power at an alpha level of 0.05. The mean age of the children was 9.85 years, (*SD* = 1.96 years) and their ages ranged from 5 years and 2 months to 13 years and 5 months.

The study received ethical approval from the School of Psychology Research Ethics Committee, Trinity College Dublin (Approval no. SPREC072023-01). To recruit the children the principal and relevant classroom teachers from each school distributed information sheets and consent forms to parents and guardians who were given seven days to provide their written consent. Each child also assented to participate in advance of the experiment. The parents or guardians reported that all children had normal or corrected-to-normal vision and hearing at the time of testing.

Stimuli and apparatus

The object stimuli included 30 images of familiar animals and their sounds. An initial selection of 9 sounds and 9 visual images of objects used in our Experiment was obtained from the Multimodal Stimulus Set provided by Schneider, Engel & Debener

(2008). The remaining stimuli, 21 visual images of animals and 21 animal sound clips, were sourced from various other online repositories (e.g. Pixabay, iStockphoto, Freesoundslibrary) and were selected to be as similar in format as the initial 18 stimuli. During the experiment, all object images were presented in the centre of the screen against a white background. To achieve this, each animal image was edited using Gimp3.1 by deleting the original background (i.e. grass or branches). The size of each animal image was also standardised to fit within a square area with a maximum width and height of 378 pixels (i.e. 6.87 cm) on a laptop screen. Once the participant was seated, each object image subtended a visual angle of approximately 6° in both the horizontal and vertical dimensions.

All sound clips were processed using the software package 'Audacity 3.1.3' to be a standard duration of 2 seconds. The sampling rate was 44.1kHz and sounds were compressed to a mono-track. During the main experiment, the sound stimuli were delivered via Sony MDR ZX310APPB wired headphones. Sound intensity was adjusted for each child prior to the main experiment: a sample sound stimulus was presented and its loudness adjusted to a level that was comfortable for the child, as required. The average intensity of the auditory stimuli was approximately 50 dB.

To validate the object stimulus selection, the 30 animal images and their respective animal sounds (30) were tested in two separate pilot studies. The first, conducted on 5 naive adults, tested participants' ability to categorise (as 'wild' or 'farm') and then identify (i.e. name) and each animal image and each animal sound stimulus. All visual stimuli were correctly categorised (i.e. 100%) and the average categorisation accuracy for the animal sounds was 94.7%. Identification rates for each of the animal images or sounds in the final selection of stimuli were over 80%. We also assessed the familiarity of the animal images and sounds in a group of 22 children aged 10-12 years in a separate study. Although children rated the 'farm' animals as more familiar than 'wild' animals ($t(21) = -2.73$, $p = 0.01$),

importantly, we found no difference in familiarity ratings between the sounds and images of the animal stimuli used in our study ($t(18) = 1.23$, $p = 0.24$).

The 'context' stimuli were created by adapting two different images of scenes, one of a 'wild' (i.e. jungle) scene and one of a 'farm' (i.e. farmyard barn) scene. Each original scene image was designed using Gimp3.1 with components sourced from Canva, an online graphic design platform. These contextual stimuli were designed in order to control visual complexity and extraneous category cue with previous research in adults finding drawings of contextual scenes evoke similar patterns of perceptual processing (Singer et al., 2023) and superordinate categorisation representation (Yao et al., 2025) as photorealistic contexts. Each image was edited to be presented as a ‘frame’ around the object stimulus in the experimental trials. Once the two context stimuli were created, these images were then converted to greyscale to ensure that colour was not used as a diagnostic cue for category membership (see Chen & Cheung, 2019; Olivia & Schyns, 1996). The context stimuli were then scaled to fit within the following dimensions: 1368 (width) x 1094 (height) pixels (i.e. 24.87 by 19.87 cm) on a laptop monitor. During the experiment, the context stimuli subtended a visual angle of approximately 21° horizontally and 12° vertically. To ensure that the same two context stimuli were not shown repeatedly in the experiment, we created 7 different versions of each of 'wild' and 'farm' context stimulus using Gimp3.1. This was achieved by randomly removing different features (representing 35% of the content) from each original scene. To create the stimuli for the experimental trials, each 'wild' or 'farm' scene context was paired with an animal image in a manner that was semantically congruent (e.g. a lion and jungle scene) or incongruent (e.g. a lion and a farmyard scene). We also presented images of objects without context (i.e. absent context).

Finally, as a visual aid for the children to respond, each trial included an image of a 'zookeeper' and 'farmer' (each character was presented from the waist upwards) which were

sourced online (via clipart-library.com), and edited to reduce their size to fit within maximum dimensions of 359 width and 300 height in pixels (i.e. 6.52cm width by 5.45cm height) on the laptop screen. The image of each character was positioned in the bottom left and right corners respectively of the screen and were presented from the start of a trial (i.e. with the object stimulus). An example of a visual display from a trial is shown in Figure 1.

Figure 1

An example of a stimulus display taken from a trial in the main experiment

Note. The animal image is presented in a central position and surrounded by a scene context (congruent in this example). The 'zookeeper' and 'farmer' characters shown at the bottom of the image are included as prompts for responding. See text for further details.

We used a laptop (Alienware 15 R4) with a 40cm (resolution of 1920 x 1080 pixels), 60Hz IPS display screen to conduct the experiment. This laptop was positioned on a desk in front of the seated participant at a distance of approximately 60cm away. This testing apparatus was set up in a quiet corner of a classroom or corridor of the hosting school. The experiment was programmed to present stimuli and collect response data using Psychopy 2023.2.3. The children wore headphones and used a mouse to input responses. If any child

did not know how to use a mouse, they responded verbally by saying ‘wild’ or ‘zoo’ and the experimenter inputted their responses[1].

Design

The experiment was based on a fully factorial, mixed design with one between and two within subjects’ factors. The between-groups factor was age which ranged continuously from 5 to 13 years. There were two within-groups factors. The first was object modality which included 3 levels: audiovisual (AV), visual alone (V), auditory alone (A). The second factor was the background context (i.e. scene) with 3 levels: congruent, incongruent or no context. The dependent variables were categorisation accuracy performance and response times.

According to the factorial design, there were 9 experimental conditions (i.e. 3 object modalities by 3 contexts). For any one participant, the 15 'wild' and 15 'farm' animal stimuli were allocated to the different conditions using a Latin-square design, organised such that there were 3 'wild' and 3 'farm' objects per condition, yielding 54 unique trials. The allocation of object stimuli to experimental condition was fully counterbalanced across participants. Each object stimulus was presented twice in the experiment, but in different experimental conditions. In addition, different versions of the context image (i.e. different versions of the 'farmyard' scene) were always presented across repeated trials. The specific version of each scene context was randomly allocated across trials and participants. The trials were presented across two different blocks, with repeated objects presented in the second block (and in different conditions). Participants could take a self-timed break between blocks.

[1] This occurred for 3 of the youngest participants only (aged 5-6 years). These participants data is included in the analysis of accuracy; however, they are excluded from our analysis of reaction times.

The experimental trials were based on the following structure: the trial started with a fixation cross for 250ms, which was immediately followed by the visual, auditory or audio-visual object stimulus, shown in a context condition, for a duration of 2000ms. The response cues (i.e. images of the farmer and zookeeper) also appeared with the object stimulus display but remained on the screen until the participant made a response. See Figure 1 for an example. Participants could respond to the trial as soon as the object stimulus appeared and were given unlimited time to respond. If more than 2000ms had elapsed without a response, the child was presented with a visual prompt in green text on the screen to remind them that the zookeeper represented the 'wild' animals ("Zookeeper = Wild") and the farmer represented the 'farm' animals ("Farmer = Farm").

Procedure

Each child was invited to sit at the table to take part in the experiment, and the laptop screen was positioned approximately 60cm away from the child. They first read written instructions on the screen, which the experimenter also repeated. The experimenter then demonstrated how to use the computer mouse, which was positioned in front of the child, to respond to each trial. Each child then completed two practice trials, one in which only an image of an animal was presented and one in which only the animal sound was presented. These practice trials could be repeated if the task instructions were still unclear, and the experimenter ensured the children understood the task before embarking on the main experiment.

The children were provided with a narrative setting in order to encourage them to complete the task. The were told that the task was a type of game and that both a farmer's animals and animals from the zoo had escaped into Phoenix Park (a large, well-known park in Dublin within which Dublin Zoo is located). The task for the child was to help the farmer and the zookeeper return each animal to their correct home by quickly categorising each

animal as a 'farm' or 'wild'. In order to respond to a trial, for each animal they saw or heard, or sometimes both see and hear, the participants were instructed to move the mouse to click on the character to whom the animal should be returned: i.e. the farmer for the category of 'farm' animals or zookeeper for the category of 'wild' animals (see Figure 1). The experimenter sat next to the participant throughout the task and, if necessary, prompted the child to pay attention to the screen during the experiment. To further ensure that children were attending to the task we also included a random number of 'catch' trials in the second last block of trials. In these trials, an image of a fly appeared on the screen, and the children were required to 'swat' it by clicking on its image as quickly as possible. We used these responses to apply an attentional cut off as follows: data from participants who responded slower than 2sd from the mean response time of age matched participants would be excluded from further analysis. On average, the experiment took approximately 10-15 minutes for each child to complete.

Results

Of the 181 participants recruited the data from 8 children were excluded from analyses resulting in a dataset from 173 participants. Of the excluded data, data from 5 participants were excluded due to poor overall performance (mean age = 10.76 years; 4 boys) and data from a further 3 participants were excluded (mean age = 12.56 years; all boys) because of a technical error with the delivery of the sound stimuli. We also analysed children's responses to the attentional check task. Their mean RT of 2.15 seconds (*SD* = 1.39) to the catch trials was within the expected range for their age: a Spearman rank-order test also revealed a positive correlation between response times to the catch trials and overall mean reaction time [$r_s(173) = 0.51$, $p < 0.001$]. Taken together, these results suggest that all children attended to the task, and it was not necessary to exclude any data. Children's overall performance to each of the object modalities and context are presented in Table 1.

Table 1

Children's mean performance across experimental conditions: object modality and semantic context of the background scene

Modality	Visual-only			Auditory-only			Audio-visual		
Context	Con	Incon	None	Con	Incon	None	Con	Incon	None
Accuracy	0.91	0.91	0.91	0.83	0.69	0.80	0.92	0.91	0.92
RT (s)	2.17	2.24	2.19	2.66	2.90	2.91	2.23	2.21	2.18

Note. Mean performance is represented by proportion accuracy (Accuracy) and response times (RT) measured in seconds; object modality (Visual-only, Auditory-only and Audiovisual) and semantic context of the background scene (congruent (con), incongruent (incon) and no context (none).

Categorisation Performance: Accuracy

The average performance of each child across all ages for each object modality (Visual-only, Auditory-only, Audio-Visual) is shown in Figure 2. A linear mixed-effects model was conducted to examine the effects of object modality, scene context (none, congruent, incongruent), and age (as a continuous predictor) on categorisation accuracy. The model included fixed effects for all main predictors and their interactions, as well as a random intercept for participant to account for individual differences. P-values and degrees of freedom were estimated using the Satterthwaite approximation.

There was a main effect of age [$F(1, 1206.6) = 6.51$, $p = .011$, partial $\eta^2 = .11$], with improved categorisation accuracy with increasing age ($\beta = 0.013$, $SE = 0.005$;[$t(1207) = 2.55$, $p < 0.01$]). Estimated slopes calculated using the emtrends package revealed that this age-related improvement was present across all modalities, although it was strongest in the visual-only ($\beta = 0.0142$) and auditory-only ($\beta = 0.0136$) modalities, and weakest in the audio-visual modality ($\beta = 0.0092$) see Figure 2 for details.

The model revealed a significant main effect of object modality [$F(2, 1368) = 4.48$, $p = .011$, partial $\eta^2 = .02$]. Post hoc pairwise comparisons were conducted with Tukey adjustment for multiple comparisons and showed that participants performed significantly more accurately in the audio-visual modality ($\beta = 0.19$, $SE = 0.06$, $t(1368) = -18.719$, $p <$

0.001), than the auditory-only modality. Performance to the visual-only modality was also significantly better than to the auditory-only modality ($\beta = 0.07$, $p = .293$, $t(1368) = -18.719$, $p < 0.001$) but did not differ from the audiovisual modality ($t(1368) = -0.321$, $p = 0.94$). The main effect of context on accuracy performance failed to reach significance, [$F(2, 1368) = 2.59$, $p = .075$, partial $\eta^2 < .001$].

There was no evidence of any two-way interaction between the factors: object modality and context [$F(4, 1368) = 2.25$, $p = .062$, partial $\eta^2 = .007$]; object modality and age, [$F(2, 560) = 0.39$, $p = .676$, $\eta^2 = .001$] context and age [$F(2, 560) = 2.03$, $p = .132$, $\eta^2 = .007$]. Similarly, the three-way interaction between object modality, context and age failed to reach significance [$F(4, 560) = 0.31$, $p = .872$, $\eta^2 = .002$].

The model explained a moderate proportion of variance in accuracy, with a marginal R^2 (fixed effects only) of .259 and a conditional R^2 (fixed and random effects) of .396. To evaluate the relative contribution of each predictor (object modality, context and age) to categorization accuracy, a series of nested mixed-effects models were compared against a baseline model with only random intercepts for participants. Each model included one fixed effect at a time, with this then compared to the whole model including all main effects and their interactions.

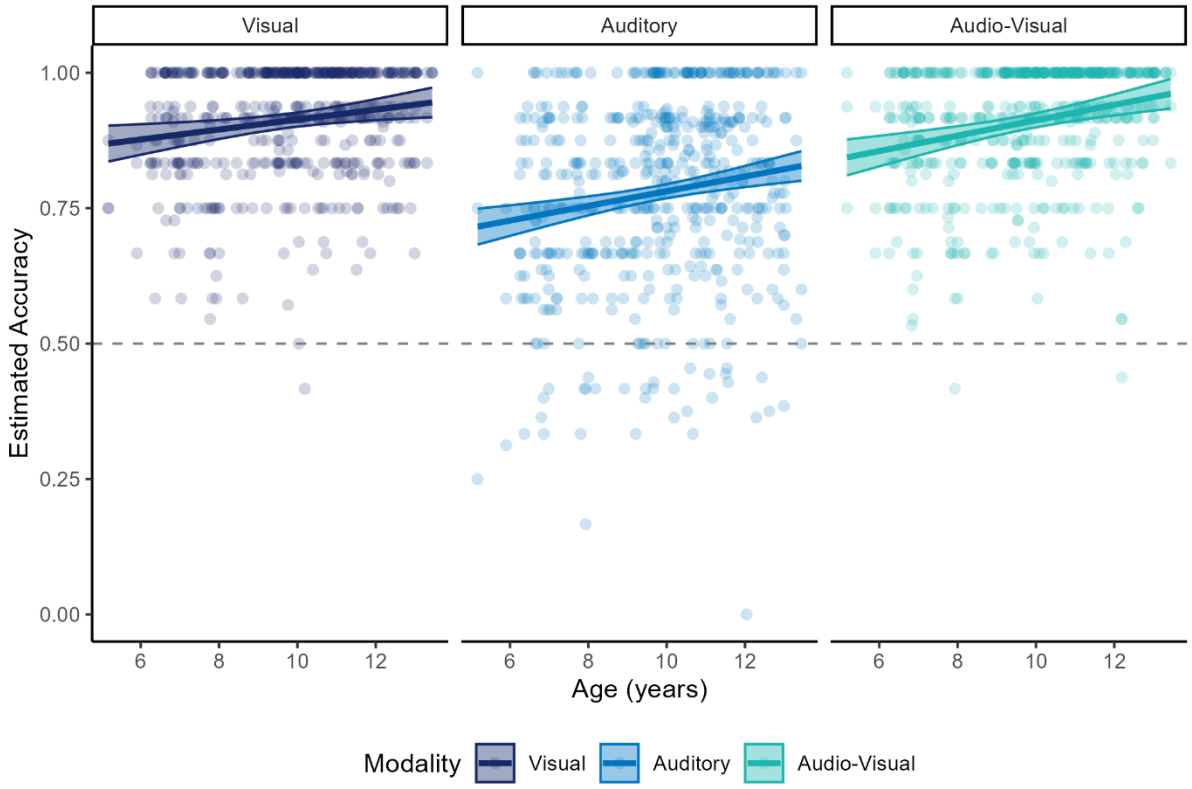
Model comparisons based on AIC, BIC, and likelihood ratio tests showed that each predictor significantly improved model fit over the null model. The model including only 'object modality' explained significantly more variance than the null model [$\chi^2(2) = 367.91$, $p < .001$]. Similarly, models including only 'scene context' [$\chi^2(2) = 37.47$, $p < .001$], and only 'age' [$\chi^2(1) = 21.04$, $p < .001$] as predictors also show significant improvements over the null model.

Further comparisons of the full model to the three single-predictor models revealed that the full model significantly outperformed the object modality-only [$\chi^2(15) = 156.31$, p

< .001], scene context-only [$\chi^2(15) = 486.75$, $p < .001$], and age-only [$\chi^2(16) = 503.18$, $p < .001$] models. This indicates that all three predictors, and their interactions, contribute uniquely to explaining the variance in performance. Model-level variance explained was examined using marginal and conditional R^2 values with the 'object modality'-only model accounting for the largest proportion of marginal variance of the single-predictor models ($R^2m = .18$), followed by age ($R^2m = .03$) and context ($R^2m = .02$). The full model explained substantially more variance ($R^2m = .26$), indicating the presence of interaction effects across the predictors which better capture their influence on categorisation accuracy - though in our analysis we did not detect any significant interactions. Model selection statistics further supported the full model as the best fitting, with the lowest AIC (–2113.6) and BIC (–2006.6) of all candidate models. Among the individual predictors, 'object modality' contributed the most substantial improvement in model fit (AIC = –1987.3), followed by 'context' (AIC = –1656.8), then age (AIC = –1642.4).

Figure 2

Plot showing categorisation accuracy across the participants according to their age



Note. Each plot represents the performance to one of the object modality conditions: Visual-only (left), auditory-only (middle) and audio-visual (right). The shaded area represents 95% confidence intervals. The dots represent each individual child's mean accuracy within each of the object modalities.

Categorisation Performance: Reaction Times

Data from trials in which the participant made anticipatory categorisation responses faster than 0.5s were first removed prior to analysis. In addition, trials in which response times were ±3 standard deviations from each participant's mean reaction time were also removed. These processes resulted in the exclusion of 2.92% of the entire data set from further analysis. The remaining reaction time data was negatively skewed; we conducted a Kolmogorov Smirnov test on mean reaction time across participants and found the

distribution of reaction times significantly deviated from normal [D = 0.16, p <0.01], therefore we logarithmically transformed the response time data.

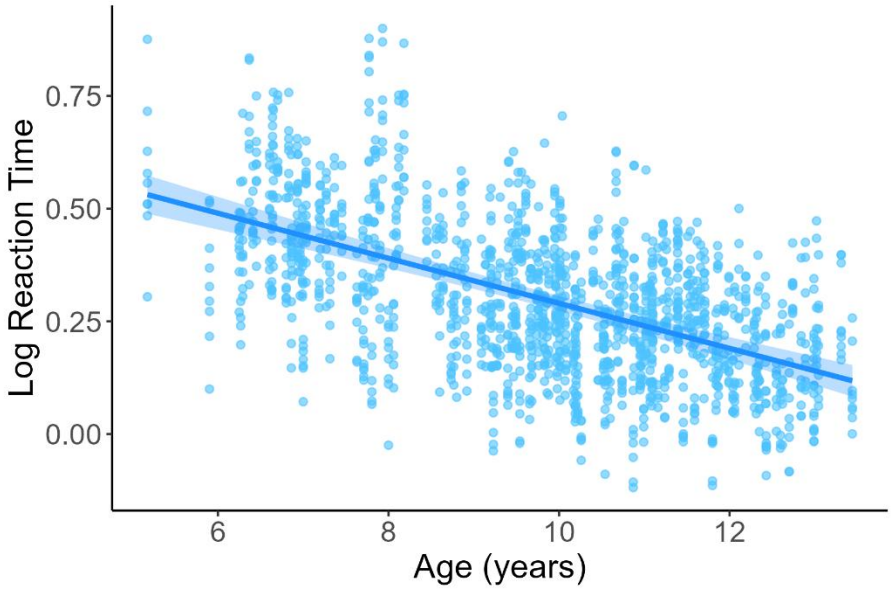
To examine the effects of object modality (auditory-only, visual-only, and audiovisual), scene context (congruent, incongruent or no context), and age (continuous) on reaction times (RTs) during the categorization task, a linear mixed-effects model was fitted to the log-transformed mean RTs. The model included the fixed effects of object modality, context, age, and all interactions, with a random intercept of 'participant' which would account for individual differences across repeated measures. Models were fitted using restricted maximum likelihood (REML). T-tests for fixed effects were computed using Satterthwaite's approximation for degrees of freedom, and F-tests for type III ANOVA were calculated using Kenward-Roger approximations, which improve inference in unbalanced designs (i.e. fewer participants in the younger than older ages). Effect sizes for fixed effects were reported as partial eta squared, and model fit was evaluated using marginal and conditional R^2, representing variance explained by fixed effects alone and by the full model including random effects, respectively. Likelihood ratio tests confirmed that including object modality, scene context, age, and their interactions significantly improved model fit over nested models. The inclusion of 'object modality' over the null model yielded $\chi^2(2) = 612.52$, $p < .001$; 'scene context' condition over the null model was $\chi^2(2) = 3.93$, $p = .140$; 'age' over the null model was $\chi^2(1) = 104.76$, $p < .001$. The full model including object modality, scene context, age, and all interactions improved the model fit over the 'object modality'-only model, $\chi^2(15) = 185.35$, $p < .001$, over the 'scene context'-only model, $\chi^2(15) = 793.94$, $p < .001$, and over the 'age'-only model, $\chi^2(16) = 693.11$, $p < .001$. Age-related slopes per modality were estimated using estimated marginal trends (emmeans), with Tukey-adjusted pairwise comparisons among estimated marginal means. Residual diagnostics indicated acceptable model assumptions, and variance inflation factors

suggested no problematic multicollinearity among predictors. The model showed a marginal R^2 of 0.426 and the conditional R^2 was 0.812, showing that the full model, including random effects, accounted for 81.2% of the variance. Residual diagnostics indicated no major violations of assumptions.

Type-III Wald *F* tests (Kenward–Roger df) revealed a main effect of age on logRT [$F(1, 171.01) = 134.83$, $p < .001$, $\eta^2_p = .44$]. Thus, as age increased, log RTs decreased (i.e. faster response times in older children; $b = -0.05$, SE = 0.0043, 95% CI [−0.06, −0.04]) as shown in Figure 3.

Figure 3

Plot showing mean categorisation response times (logarithmically transformed) across all participants



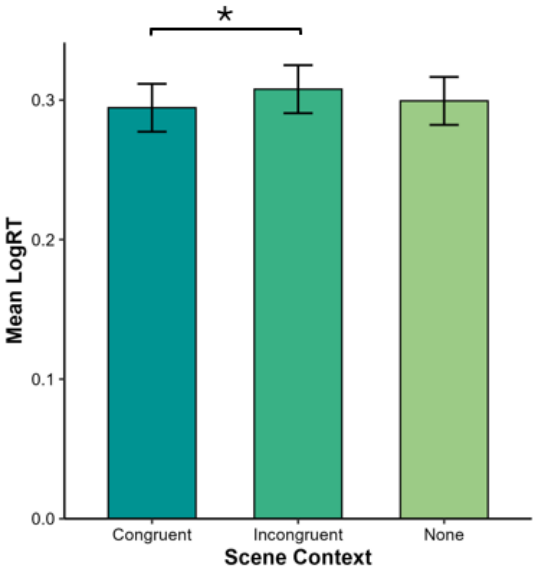
Note. The shaded area represents 95% confidence intervals. The dots represent each child's mean reaction time across age in years (continuous).

The main effect of modality was not significant $F(2, 1343.02) = 1.53$, $p = .22$, $\eta^2_p = .002$]. There was a main effect of scene context [$F(2, 1343.02) = 7.83$, $p < .001$, $\eta^2_p = .01$] as shown in Figure 4, with significantly longer responses (logRT) when the scene context was incongruent ($M = 0.31$, SE = 0.009, 95% CI [0.29, 0.32]) relative to a congruent

scene context condition (M = 0.29, SE = 0.009, 95% CI [0.27, 0.31]; t(1343) = 2.92, p = .01). Although response times to the 'no context' (M = 0.29, SE = 0.009; 95% CI [0.28, 0.32]) were slower than those to the congruent condition, and faster than those to the incongruent condition, none of the pairwise comparisons were significant (ps > .05).

Figure 4

Plot showing categorisation mean reaction times (logarithmically transformed) across scene context conditions

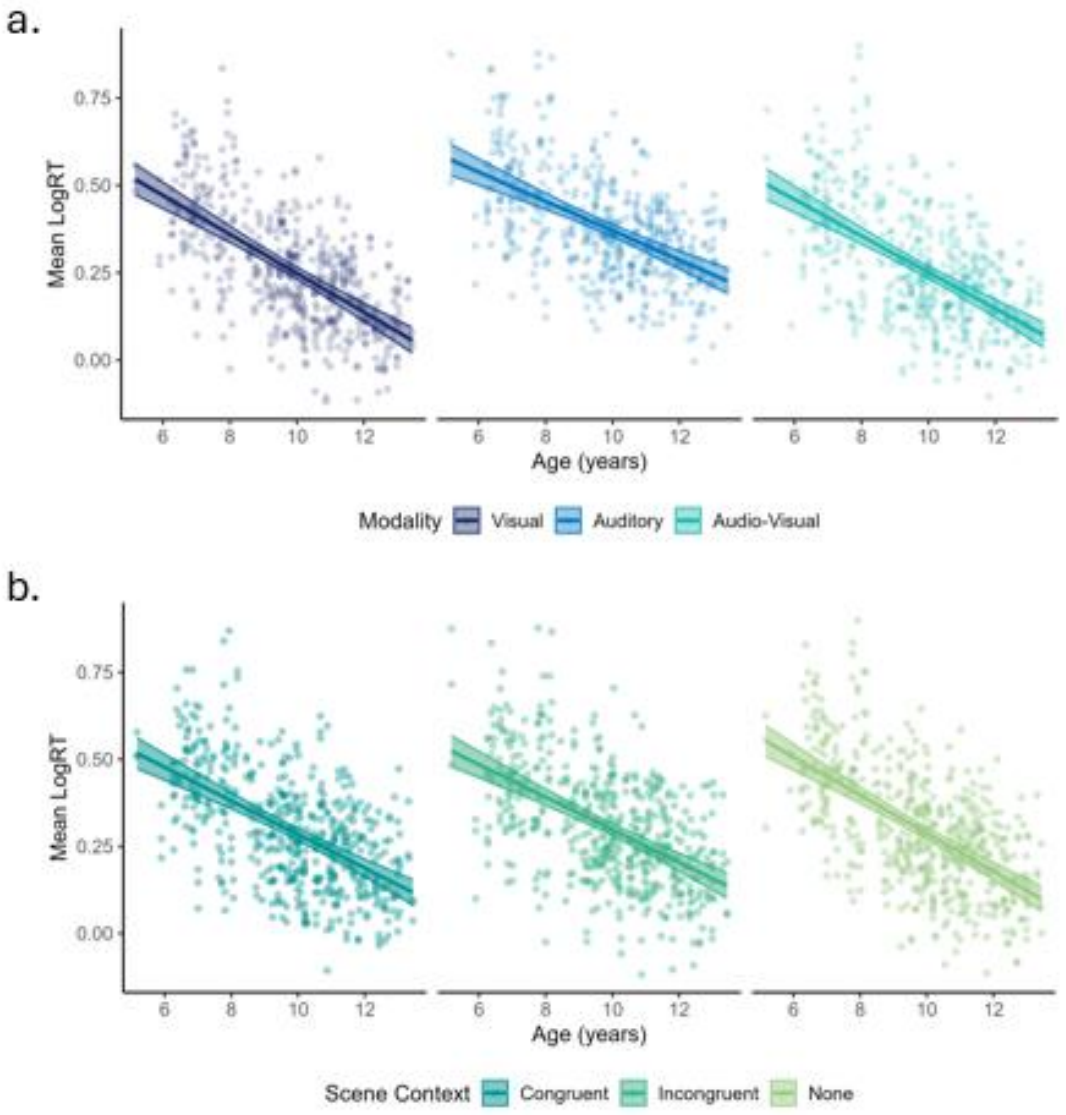


Note. The error bars represent 95% confidence intervals; * = $p < 0.05$.

This effect of object modality was qualified by a significant interaction with age [F(2, 1343.02) = 22.49, p < 0.001, η^2_p =.03] as shown in Figure 5a. Post hoc analysis revealed that response times (logRTs) decreased with age across all modalities, but this decline was steepest in the Audio-Visual (b = -0.0065, SE = 0.005, 95% CI −0.067, −0.048], p < .001) and Visual-only (b = -0.0057, SE = 0.005, 95% CI [−0.0061, −0.0044], p < 0.001) modalities relative to the slope of response times in the Auditory-only modality (b = -0.0041, SE = 0.0047, 95% CI [−0.0051, 0.0033], p < .001) which did not significantly improve with age. The slopes to the AV and V-only conditions did not differ from each other (p > .19).

Figure 5

Plots showing mean reaction times (logarithmically transformed) across participant ages for a) each of the object modality conditions and b) each of the context conditions



Note. The individual dots represent the mean logRTs for each individual child. The shaded area represents 95% CIs.

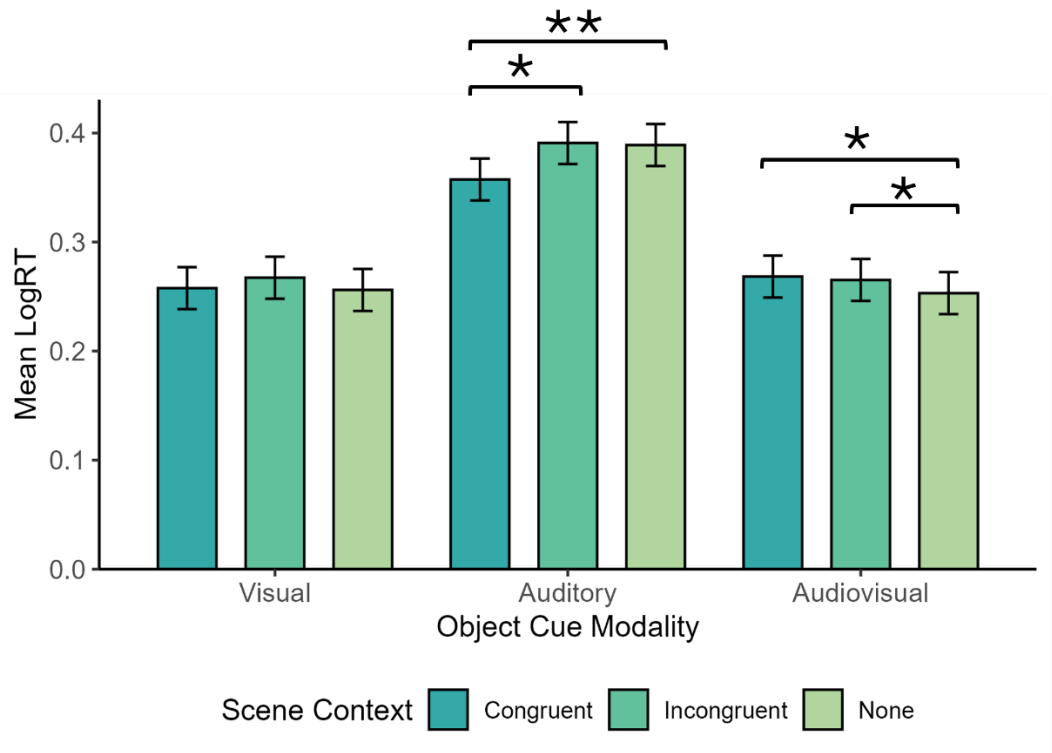
There was evidence for a 2-way interaction between scene context and age [$F(2, 1343.02) = 8.80$, $p < 0.001$, partial $\eta^2 = .01$] which is shown in Figure 5b. A comparison of the slopes suggested that age-related improvements were comparable across the congruent (−0.049, SE = 0.005, 95% CI [−0.058, −0.040]) and incongruent scene contexts (−0.047, SE = 0.005, 95% CI [−0.056, −0.038]), [$t(1343) = -1.01$, $p = .57$,

r < .01]. However, the slope to the 'no context' condition (labelled 'none' in Figure 5b) was significantly steeper (−0.056, SE = 0.005, 95% CI [−0.066, −0.047]) than that to either the 'congruent', [$t(1343) = 3.02$, $p = .007$, r = .08] and 'incongruent', [$t(1343) = 4.03$, $p = .002$, r = .08] conditions.

The pairwise interaction between scene context and modality [$F(4, 1434.02) = 3.61$, $p = .006$, $\eta^2_p = .010$,] was significant, and is shown in Figure 6. A congruent context resulted in faster RTs to auditory objects relative to an incongruent ($t(1343) = -4.04$, $p < 0.001$) or no ($t(1343) = -4.01$, $p < 0.001$) context. No other significant pairwise differences were found.

Figure 6

Plot showing the children's mean (log transformed) reaction times across the different experimental conditions of object modality and contextual cues (congruent, incongruent or no context)



Note. The error bars represent 95% confidence intervals; * = p < 0.05. ** = p < 0.01.

The three-way interaction between object cue modality, contextual congruence, and age did not reach significance [$F(4, 1343.02) = 2.47$, $p = .058$, partial $\eta^2 = .007$].

Discussion

The current study examined the ability of children, aged from 5 to 13 years of age, to categorise objects based on multisensory object cues and background scene context. Specifically, we investigated whether a combination of auditory and visual object information (e.g. an image of a lion with the sound of roaring) influenced children's ability to categorise objects relative to each unisensory cue alone. We also investigated whether a semantically relevant scene context further influenced their categorisation performance.

First, we found an overall improvement in categorisation performance with increasing age in terms of both accuracy and reaction time, which is consistent with many previous reports (Bjorklund, 1985; Broadbent & White et al., 2018). We also found an effect of modality with reduced accuracy to auditory-only object cues relative to either visual-only or audiovisual object cues for categorisation. In contrast to expectations based on previous reports (Broadbent et al., 2020), our results suggest no specific advantage for multisensory over visual-only object cues on either categorisation accuracy or response time performance. However, children clearly find it difficult to categorise object sounds alone.

We noticed that, overall, children had little difficulty with the categorisation task. The high accuracy performance to the visual categorisation task may be driven by children's frequent exposure to animal representations through literature, entertainment and school material (consider the popularity of the songs such as ‘Old McDonald’ or films concerning animals targeted towards children aged between 5-13 years). In contrast, there are several possible reasons to explain children's relatively poor categorisation of animal sounds. First, there may be differences in the rate at which the sensory modalities develop (see e.g.

Hendrickson et al., 2019; Robinson & Sloutsky, 2004a; Litovsky, 2015; Krishnan et al., 2013) with visual object perception more mature than auditory object perception. The underlying category structure between object sounds and object images may also differ such that images may be more discriminable than their corresponding sounds. There could also be an influence of the relative familiarity of the object information (Tanaka et al., 2005; Anaki & Bentin, 2009): e.g. an image of a lion may be relatively more familiar to a child than the sound of its roar. Regarding the latter, we did ensure that the images and sounds were equally familiar (see above), at least to older children. Future studies in which item discriminability or familiarity within and across modalities are controlled may help provide further insights into the specific role of development.

The findings concerning performance across the object modalities do not align with our initial hypothesis, or previous reports (Nardini et al., 2016), that audiovisual exposure would lead to enhanced performance relative to the visual-only condition especially in older children. Performance in both the vision-only and audiovisual conditions was highly accurate overall, particularly in older children (i.e. over 90%), suggestive of visual dominance. However, this high level of performance in older children also suggests that the tasks involving vision were relatively easy, therefore further improvement in categorisation accuracy with the addition of audition was less likely to be observed due to ceiling effects. Indeed, a benefit of multisensory inputs may be more difficult to observe when performance to any unisensory input is already high due to inverse effectiveness (Stein & Meredith, 1993). A multisensory benefit is most likely when the unisensory information is uncertain (Heron et al., 2004) or perhaps in situations of high cognitive load (see Broadbent, Osburn and colleagues (2018). It is also worth noting that the object sounds and (static) images were associated at a semantic level (Heikkilä & Tiippana, 2016) and may not necessarily tap into the perceptual processes involved in multisensory integration

(such as causal inference). For practical reasons, it was beyond the scope of the current experiment to test these crossmodal interactions through further experimental manipulations (such as adding visual 'noise' to reduce accuracy or including a condition with category incongruent sounds).

Regarding the scene context we predicted that object categorisation would be improved if presented with a congruent scene context (i.e. a lion presented with a jungle scene) relative to incongruent or no context as suggested from previous findings with adults (Roberts et al., 2024; Murphy et al., 1989) and children (Broadbent, White et al., 2018). However, this prediction was not supported by the results on accuracy. It is possible that the cognitive effort involved in children's processing of the context, due to task switching (Peng et al., 2018), may have reduced any benefit of congruent information on accuracy performance. However, a congruent context did lead to faster categorisation responses than an incongruent context. Furthermore, categorisation times to the 'no context' condition were slower than those to the congruent context and faster than the incongruent context as we expected, but these differences failed to reach significance. Given that the benefit on response time to congruent scene contexts was relative only to scenes that were semantically unrelated to the object, we are cautious about describing congruency as having a general advantage on categorisation performance. However, when only the sound of the object was presented, we found a distinct advantage on categorisation times for the presence of a semantically congruent scene relative to no context. For this modality only, no difference was observed between the 'no' context and the 'incongruent' context, suggesting that the congruent cue facilitated categorisation performance. Thus, categorisation responses were faster to animal sounds presented with a semantically relevant visual background than any other scene type. It is important to note that it is unlikely this result is due to the children simply categorising the scene alone, since accuracy

performance was not affected by the context. Instead, the benefit of the semantic scene context on sound categorisation is consistent with previous findings with adults (Shafiro et al., 2016)

Interestingly, children categorising audio-visual objects did not benefit from an additional scene context cue. This suggests that performance in the multisensory condition may have been dependent on children's attentional resources, with the presence of any scene background adding additional demands to these resources (see e.g. Goldenberg & Johnson, 2015; Sloutsky & Robinson, 2008). This was similarly observed, when a congruent context was presented with a visual-only object cue, there was no benefit of scene. Again, this is consistent with attentional demands and the relative competition between two visual cues (Porcu et al., 2014). For example, when there is a demand on resources, a prioritisation of visual information across the visual field (Krishna et al., 2017; Ecker et al., 2013) may occur such that the centrally located object cue might dominate the perceptual judgement. To address this issue of competing resources, it would be interesting to expand the current and previous work (e.g. Broadbent et al., 2020) to include both visual and auditory scene information in future studies.

Categorisation accuracy to any of the modality cues was not influenced by the other factors, including context. However, response times to the audio-visual and unisensory object cues were influenced by the age of the children and the type of scene context presented with the object. First, we found a difference in the trajectory of performance across unisensory versus audiovisual object cue modalities with age. Specifically, whilst categorisation times to both vision and audiovisual object cues significantly decreased with age, response times to the auditory-only object cue did not improve with age (i.e. there was no significant reduction to reaction times observed within our age range). This difference may be related to children's ability to efficiently process object information across different

modalities, with objects sounds being the most difficult (Berland et al., 2015; Barrett & Newell, 2015). On the other hand, this result may be related to the specific requirements of our task and inherent differences in the spatio-temporal nature of image and sound stimuli. For example, object information is readily available in visual-only presentations, therefore response times in this condition are constrained by motor development only. In contrast, sound stimuli require time to be presented: in our experiment an animal sound was played during a 2 second timeframe. Overall, we observed that children's response times were over 2 seconds long which was ample time for the exposure of the auditory information. Nevertheless, irrespective of the child's age the same minimum exposure to the auditory information was required in order to recognise and categorise the objects. Given this requirement for the sound stimuli only, it may not be surprising that the speed at which objects are categorised by sound does not improve with age. One suggestion for future investigations is to render auditory and visual information more equivalent. This could be achieved by manipulating the amount of visual information available over the same temporal duration as the sound stimulus (see e.g. Williams et al., 2022).

Scene perception, or gist, has been shown to occur at the preattentional stage of processing (Watson et al., 2025), it is both rapid (Thorpe, et al., 1996) and efficient (McCotter et al., 2005). However, evidence supporting a benefit of semantic congruency of scenes on object perception is mixed. On the one hand, there is evidence to support the idea that semantic congruence facilitates object recognition relative to incongruent scenes (e.g. Palmer, 1975; Biedermann et al., 1982). On the other hand, Wille & Ebersbach (2016) report that semantic incongruency may lead to ‘pop out’ effects which can also facilitate recognition. Thus, the unexpectedness of seeing a lion in a farm scene may influence a more rapid response than seeing a lion in a jungle. Furthermore, the results from Murphy & Wisniewski (1989) suggest that objects presented in a congruent scene encourages more

superordinate category judgements (e.g. 'wild animal') whereas objects presented in incongruent scenes lead to relatively faster, basic level category judgements (i.e. 'a lion'), at least in adults). Whether a benefit of congruent over incongruent scenes can be attributed to the speed at which children can categorise objects at a superordinate or basic level is currently unclear without further research. Though the usage of an animated scene context in grayscale was selected due to the ability to control for low level salience and visual complexity this may have further affected categorisation performance. Research suggests the usage of drawn images supports superordinate categorisation as opposed to categorisation at the basic level (Yao et al., 2025), in order to clarify this potential source of ambiguity future work may consider the use of photorealistic contextual stimuli.

The results of this study investigating the ability of children to categorise familiar stimuli from the age of 5 years to 13 years confirms that categorisation accuracy and time improves with age. Relative to the body of literature concerning category formation in both infants and adults, investigations of categorisation in this age group are understudied, particularly their capacity to incorporate multisensory as well as contextual scene information into their category judgements. Although children found that object sounds were difficult to categorise, children's ability to use sounds for categorisation is known to be worse than that of adults (Aslin & Smith, 1988; Roark et al., 2023). The relative influence of associated scene context on performance was mainly limited to categorisation times, and there was no evidence that children's categorisation accuracy benefitted from this semantically relevant cue. In fact, unlike the results from Broadbent, Osborne et al., (2020) which suggested a benefit for two cues over one in the formation of novel categories, our results suggest that children’s categorisation of familiar objects does not always benefit from extra information, suggesting that their ability to categorise objects based on multisensory information is constrained by attentional resources across development.

Chapter 2

Chapter 3

Does sound improve the categorisation of moving objects in children and adults?

Chapter 3

Abstract

Children's ability to categorise objects based on visual shape emerges early, but the impact of additional information on the categorisation process is unclear. This study examined how audiovisual motion cues influence category learning and subsequent object categorisation across development. Three different age groups of children (5-8yrs; 9-10yrs; 11-13yrs) and one adult group first learned to categorise novel moving, objects presented with sounds to a learning criterion. Categorisation was then tested across all age groups using the same shaped objects that were learned. During the test, these objects were either presented with the same motion and sound as learning, or with a different motion and sound, or were shown as static images (with original or different sound). All age groups successfully acquired the novel object categories, with a more rapid learning trajectory in adults than children. At test, the adults outperformed the children overall, but there was no difference across the age groups of the children. For children and adults, performance was best for the learned object and movement pairings and a change in object motion reduced categorisation accuracy. However, children's categorisation performance particularly benefitted from the presentation of the learned object sound with the learned object motion. Our results suggest that multiple object cues influence children's ability to learn novel categories of objects. These findings also highlight the role of crossmodal information in category formation across development.

Chapter 3

Introduction

An object category is a representation of objects which share certain perceptual or semantic qualities while maintaining discriminably distinct elements (Bornstein & Arterberry, 2010). In a world filled with sensory stimulation, the perceptual organisation of similar objects into categories allows us to make sense of what we encounter and to generalise from known to novel exemplars (Smith & Medin, 1981) thus supporting learning and memory (Carey, 2009; Mandler, 2004; Rosch et al., 1976). Therefore, categorisation may be guided by both perceptual (Quinn & Eimas, 1996a; 1996b) and semantic experience throughout development (Mareschal & Quinn, 2001; Mareschal et al., 2003; Behl-Chadha, 1996; Pauen, 2002). The ability to organise objects into categories during childhood is particularly important due to the frequency of encounters with new objects: without the ability to categorise, each novel experience would require a unique response (Rakison & Oakes, 2003; Bornstein & Arterberry, 2010). Children's organisation of objects into distinct categories emerges early, as young as 3-4 months old (Quinn et al., 1993). Furthermore, these categories appear to be based on the perceived characteristics of objects including their shape (Graham & Poulin-Dubois, 1999; Graham & Diesendruck, 2010) as well as their movement (Poulin-Dubois et al., 1996). Despite these findings, our understanding of the association of shape and motion cues in the formation of novel object categories in children is, however, poorly understood and is the focus of the current study.

Our knowledge of the role of motion on children's ability to categorise objects comes mainly from studies of living and non-living categories. Evidence for the early emergence of perceived animacy (Hofrichter et al., 2021; Kellman et al., 2006; Spriet et al., 2022; Opfer & Gelman, 2011) even in infants as young as 7-months (Taniguchi et al., 2020; Younger & Fearing, 2000) supports the importance of motion in object categorisation during development. Indeed, biological motion (Johansson, 1973) can be perceived from early infancy (Bidet-Ildei et al. 2014; Kuhlmeier et al., 2010). Other findings suggest that

in children aged 3-years of age, movement alone can aid the recognition of animate stimuli, such as humans and animals such as birds and dogs (Pavlova et al., 2001). Neural representations in the adult brain also support a distinction between living and non-living things, as evidenced from patient studies (Caramazza & Shelton, 1998; Warrington & Shallice, 1984), and neuroimaging (e.g. Sha et al., 2015). Behavioural studies suggest that animate items are recalled and recognised more accurately than inanimate ones, known as the "animacy effect" (e.g., Nairne et al., 2013; Bonin et al., 2015; Serra et al., 2023). Interestingly, the distinction between animate and inanimate categories is also supported by research using auditory stimuli which have revealed distinct neural regions activated by sounds relating to living or non-living objects, as well as action-related sounds produced by human, animal, mechanical or environmental sources (Engel et al., 2009). The findings from developmental (Mandler, 2008; Träuble et al., 2014), neuropsychological (Warrington & Shallice, 1984; Caramazza & Shelton, 1998; Ralph et al., 2007), and cognitive research (Nairne et al., 2013; Komar et al., 2022) therefore converge to suggest that object motion, either from vision or sound, can act as an important source of information for categorising an object and may have evolutionary and biological significance.

Object motion can also provide access to other rich information about an object's properties, other than shape (Biederman, 1987), which might also be relevant for object categorisation, such as its material or weight (Bingham et al., 2018; Ujitoko et al., 2023), elasticity (Schmidt et al., 2017; Ujitoko & Kawabe, 2020), and size (Tozawa & Oyama, 2006). Infants' categorisation of moving objects is, however, dependent on their ability to perceive object motion (Poulin-Dubois et al., 1996). Previous findings suggest that the perception of object movement emerges early in development. For example, newborn infants preferentially attend to dynamic objects, including faces (Spencer et al., 2006; Bastianello et al., 2022), voices (Blasi et al., 2011), or other stimuli (Otsuka & Yumaguchi,

2003) such as moving dots (Lunghi & Di Giorgio, 2024). Infants as young as 2.5 months, can perceive moving objects as cohesive units (Aguiar & Baillargeon, 1999; Spelke, 1990; Valenza, Leo, Gava, & Simion, 2006) and 6-month-old infants are sensitive to the spatiotemporal properties of objects (Leslie & Keeble, 1987). Furthermore, infants show sensitivity to differences in how objects move (Simion et al., 2008; Bertenthal et al., 1984), such as distinguishing self-propelled motion (considered a characteristic of living things; Bertenthal, 1993), from motion caused by an external force (Baillargeon et al., 2009; Pulverman et al., 2008). By about 4 to 6 months of age, infants can detect biological motion (e.g., point-light walkers vs rigid motion) (Bertenthal, 1984; Simion, 2008) which continues to be refined with development (Sifre et al., 2018; Bogfjellmo et al., 2014; Nguyen et al., 2025). Apart from motion cues to animacy, infants can distinguish between different types of object motion such as rigid versus non-rigid movement (Bardi et al., 2008) or global versus local motion (Ghim & Eimas, 1988) which are also important for categorisation (Troje, 2013; Setti & Newell, 2010). Taken together, this evidence indicates that motion is a robust and informative cue that supports both object categorization and recognition from early childhood.

In the real world, when an object moves this movement is often associated with a sound; consider the sound of a car engine, the clip-clop of a horse's hooves or the ticking of a clock. Thus, the sound of the object movement itself may be a useful cue for object categorisation. Interestingly, young infants often preferentially attend to auditory over visual stimuli (Lewkowicz, 1988a; Robinson & Sloutsky, 2004b; 2019), suggesting that object sounds acquired early in development are highly salient, with important implications for language acquisition (Yi et al., 2016). Indeed, infants aged between 8 and 10 months specifically attend to the moving mouth of a speaker, indicative of learning the crossmodal associations underpinning the production of speech in that age group (Lewkowicz et al.

2012). Voice and lip movements are a very good example of how motion and sound cues are correlated to support speech perception, yet it is unclear how motion and sound cues interact in children's formation of object categories. The ability to form cross-modal associations using object information across modalities has been demonstrated in 5–7-month-olds (Bahrick et al., 2005; Flom et al., 2009; Hannon et al., 2017), and infants aged 6-8 months can already detect temporal synchrony across vision and audition (Lewkowicz & Lickliter, 1998; Cirelli et al., 2024). Indeed, at approximately 7 months of age infants are capable of categorising sounds based on their rhythm (Hannon & Johnson, 2005). Although audiovisual perception improves into childhood, by approximately 11 months infants can already acquire audiovisual associations to categorise novel exemplars of animals (Vukatana et al., 2020; Zepeda & Graham, 2019).

In general, children's performance in categorisation tasks is lower than that of adults (Minda & Smith, 2010; Minda et al., 2008; Huang-Polluck et al., 2011), which may be due to differences in cognitive abilities, such as attention or executive function, due to the relatively slow development of prefrontal cortical regions (Rabi & Minda, 2014; Roark et al., 2024). One such consequence is children's inability to adequately suppress irrelevant (non-diagnostic) featural information during the categorisation process (Deng & Sloutsky, 2015). Selective attention has been proposed to be a central component to successful category learning, though this is still poorly optimised in children aged 4–5-years (Sloutsky, 2016; Robinson & Sloutsky, 2013, Deng & Sloutsky, 2016). Reetzke and colleagues (2016) investigated categorisation performance to novel audiovisual associations (using Gabor patches and 'ripple' sound stimuli) in children aged 7-12 years and found a significant improvement in categorisation performance from children aged 7-12 years to adolescents (13-17 years) and from children to adults. Despite developmental

differences, evidence suggests that sounds can aid visual detection (Li & Deng, 2023) as well as incidental category learning in children (Broadbent et al., 2018).

The present study investigated whether children aged from 5 to 13 years utilise the cross-modal cues of object movement and sound when forming novel object categories. To that end, participants first learned to categorise novel objects based on a combination of their shape, motion and the sound of its motion. We then examined whether object categorisation was affected by changes to the motion or sound cues, i.e. when motion cues were incongruent, or absent, and when the sound was uninformative, relative to the learned cue combinations. We also tested an adult comparison group, to assess whether children's ability to combine or associate cues was similar to that of adults or whether there were differences in categorisation performance with age. Our aim was to address two central questions: (1) Do children use object motion and sound to support object categorisation when shape is fully informative? and (2) does the effect of motion and sound on object categorisation change with development?

Method

Participants

We initially recruited 139 children aged between 5 and 13 years to take part in this study from three primary schools (children from 1st to 6th class) within the greater Dublin area and surrounding counties in Ireland. The children were grouped according to their age comprising three groups with 5–8-year-olds, 9–10-year-olds and 11–13-year-olds. Children whose age was between these groups were organised into the same group as their class peers. The selected age range (5–13 years) captures key developmental changes in audiovisual processing, particularly the progressive narrowing of the temporal binding window (TBW), which supports more precise binding of temporally related signals across childhood with the TBW demonstrating immaturity at 10-11 years (Hillock et al., 2011;

Kaganovich, 2017). Younger children are more likely to bind temporally proximal but irrelevant audiovisual cues, whereas older children show increasing selectivity and improved use of statistically reliable multisensory information (Gori et al., 2012; Nardini et al., 2008; Nardini et al., 2016). Thus, these age groups allow examination of how sensitivity to correlated versus uncorrelated audiovisual cues may differentially influence categorisation across development. We also recruited 37 adults aged 20-32 years from within Trinity College and the wider Dublin area in order to further investigate potential age-related differences. Adult recruitment was carried out through advertisements placed across the university campus and via social media.

To recruit the children the principal and relevant classroom teachers from each school distributed information sheets and consent forms to parents and guardians who were given seven days to provide their written consent. Each child also assented to participate in advance of the experiment. The parents or guardians reported that all children had normal or corrected-to-normal vision and hearing at the time of testing. All adults provided informed consent prior to testing and reported normal or corrected-to-normal vision and normal hearing prior on the day of testing. Further demographic details of the participants can be found in Table 2.

Table 2

Demographic details of the participants across all age groups

Age Group	(n)	*M age (years)*	*SD (years)*	Min age	Max age	% Males
5-8 years	52	7.63	0.90	5.25	8.92	44
9-10 years	37	10.0	0.53	9.02	10.9	51
11-13 years	50	12.1	0.55	11	12.9	40
Adult	37	25.3	3.43	20.4	32.9	36

An a priori power analysis conducted using G*power 3.1.9.7 revealed a sample of 28 participants per age group would result in an 80% power at alpha level 0.05 to detect a

medium effect ($f = 0.25$). This was increased to a target sample of 36 participants per age group in order to counterbalance the stimuli conditions across participants (see Design section for details on the counterbalancing procedures). The study received ethical approval from the School of Psychology Research Ethics Committee, Trinity College Dublin (Approval no. SPREC072023-01).

Stimuli and apparatus

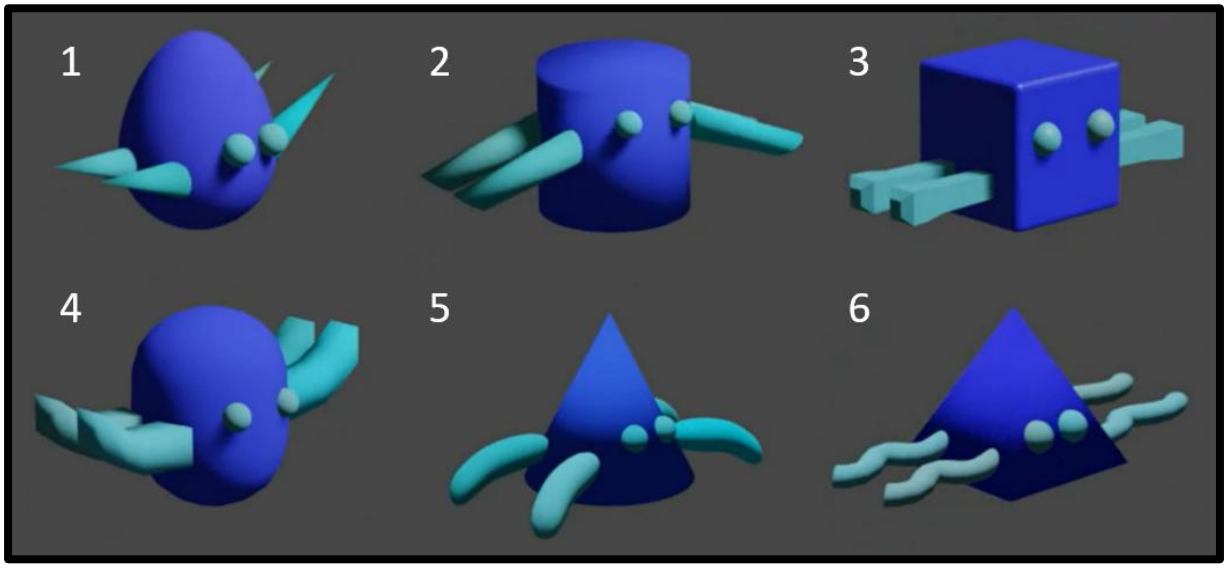
We created a set of 6 novel object shapes, using the 3D modelling software Blender 3.5.0, for use as stimuli in our experiment (see Figure 7). All objects were based on different 3D shape primitives (an ovoid, cube, square based pyramid, egg, cone and cylinder) and their configurations. The overall structure of the objects was similar such that each comprised a large central body to which four smaller parts (legs) were attached, with two parts positioned on opposite sides of the body. Both the shape of the body and parts were unique to each of the 6 objects. The object parts were also comprised of unique orientations: that is, the parts of each object all pointed either 45° upwards, 45° downwards or oriented 0° (parallel) with respect to the ground. We attached two small convex hemispheres, side by side, onto the middle of each of the objects' body. These features were identical across all objects and therefore had no experimental purpose. However, they resembled eyes which gave the objects a more 'life-like' appearance to help engage the children. Within the Blender environment, the size of the different object bodies was limited to within an approximate volume of $2cm^3$. The parts were also limited to a minimum of 0.159cm and maximum of 0.65cm in width, and a length of 1.88cm. All objects had a dark blue body, and their legs and eyes were of a different, lighter blue (see Figure 7).

Within the 3D virtual space, each object was positioned on a 20x20 cm floor shaded grey and the 'camera' was positioned using the following coordinates: 27.5 in the X axis, 26.5 in the Y axis and 14.75 in the Z axis. In effect, the camera viewed the object at an

angle of 75° in the X plane and 45° in the Z plane (see Figure S1 under Supplementary Materials). The lighting source was set to be diffuse and 1000 watts (positioned at location 0,20,20 along the X, Y, and Z axes in the virtual space).

Figure 7

An illustration of each of the six objects (1-6) used as stimuli in the experiment



Note. The objects were made using unique parts but were similarly structured with large bodies and smaller parts attached to the body.

Each of the six object stimuli were then animated to depict unique movements. For all objects, the body and legs moved together rigidly, i.e. the legs moved in tandem with the body. All six objects moved from the same starting point and ended on the same end point within the rendering environment by translating along an invisible zig-zag path (of 60° turns from left to right) along the X and Z planes (i.e. diagonally across the floor). The six distinct object movements were tumbling, twirling, turning, bouncing, rocking and pulsing. The tumbling movement involved a 360° rotation of the object in a forwards and backwards direction; the twirling movement involved slow alternating 360° turns to the right and then to the left; the turning movement involved the object turning from right to left i.e. -45° to centre and then 45° to centre; the bouncing movement involved a regular

upwards and downwards movement from the floor; the rocking movement involved a movement of the objects vertical axis 60° forwards and backwards; and the pulsing movement involved the object rapidly moving towards and away along the depth (i.e. Z) axis which gave the appearance of a ‘looming’ and 'receding' movement. Examples of these movements can be found here: [https://osf.io/d4r8w/overview?view_only=d9bf5b0a1d0548368fdbd81e700d6bef]. Each movement pattern was recorded independently of the object shape which allowed us to mix the shape-movement pairings for the purpose of our experiment (see Design). We also included images of the object without movement, i.e. static images, which were taken from the animation sequence and selected from a canonical viewpoint so that all of the object’s features were visible in the image.

The object animations were rendered as video recordings from the Blender workspace. Each video was rendered at a rate of 26 frames per second and exported to a ‘mp4’ format. Each video clip had a resolution of 1920x1080px and had a duration of 6 seconds. In the static condition, each object was positioned in the centre of the virtual space (at Z = 0) and also presented for 6 seconds. During the experiment, these video clips (and static images) were presented from a central position, against a grey background, and extended 22cm in width and 11.3 cm in height on a laptop screen. The display on the screen also included two images of planets, one in each of the bottom left (e.g. pink planet) and right (e.g. yellow planet) corners. Each image of a planet covered a $4cm^2$ area of the screen. The planets were used as an aid for the categorisation task and were each named planet ‘Bloop’ or planet ‘Neem’.

The sound stimuli were generated using soundtrap (www.soundtrap.com), an online music generation site, using the ‘Violin – Chamber’ sound effect. These stimuli were then edited using Audacity to ensure that the sounds matched with the type of object motion

depicted in each animation. To achieve this, we adopted known crossmodal correspondences between elevation and pitch (Motoki et al., 2023; Spence, 2020): changes in sound pitch matched any changes in the position of the objects in the Z and Y planes (note that all objects followed the same trajectory along the floor) during the six different animation sequences. The extent and speed of pitch increase/decrease was associated with the extent and speed of the spatial displacement, and the modulation in pitch was temporally synchronised with a change in movement direction. If the object remained in a stable position at any time during the animation, a continuous sound was heard, and the pitch remained unchanged for the same duration. As the sound pitch corresponded with the object's movement patterns, the sounds therefore 'correlated' with the motion, defined by both the direction of the object movement (higher elevation = higher pitch, and lower elevation = lower pitch) and the timepoint at which the movement changed. Since an object's shape can cause minor disruptions to its movement, when we created 'novel' object-motion pairings for the categorisation test, it was sometimes necessary to use the phase shifting function on Audacity (e.g. auditory stimulus with a 1.5s phase was shifted to match a novel 2s cycle movement pairing) to ensure that the sound was always correlated with the movement.

For the purpose of the experiment, we also wished to include a control condition in which the sound was not informative of object motion. To create these sounds, we used a Fourier transformation to filter each of the original sound clips. This filtering effectively disrupted the temporal information in the sounds but preserved the overall pitch limits to match the range of pitch included in the original (correlated) sounds. We subsequently refer to the transformed sounds as 'uncorrelated' sounds. All auditory stimuli were standardised to have a sampling rate of 44Hz and all sounds were compressed to a mono-track when exported as wav format.

Prior to the main experiment, we conducted a pilot study (N=14) to validate these sound-object motion pairings. Our results suggested that participants rated the object movement and corresponding sounds as significantly more ‘correlated’ (M = 5.49, SD = 1.00) than the object movement paired with the transformed, ‘uncorrelated’ sounds (M = 3.08, SD = 1.12). See Supplemental Materials 2 for further details.

The testing of children took place in a quiet area (or room) within their respective schools. All adult testing took place on site in a dedicated testing suite in the Institute of Neuroscience at Trinity College. The experiment was programmed and delivered using Psychopy (version 2023.2.3). The laptop model was a Lenovo Yoga Slim 7 with screen dimensions of 29cm x 18.5cm. The laptop was positioned on a desk such that the screen was at a distance of approximately 45cm away from the participant. Consequently, each object stimulus subtended a visual angle of approximately 3° in the vertical and 4° in the horizontal plane. The auditory stimuli were delivered via Sony MDR ZX310APPB wired headphones which the participant wore for the duration of the experiment. If needed, prior to the experiment, the experimenter adjusted the sound intensity until it was at a comfortable level for the participant. The average intensity of the auditory stimuli was approximately 50dB. Participants used the ‘a’ and ‘l’ keys on the laptop keyboard to input their responses, which were indicated using different coloured stickers.

Design

For each participant, each of the six object shapes were paired with one of the six motions and their correlated sound, resulting in six different object-movement pairings which were used during learning. These object-movement pairings were counterbalanced across participants using a Latin square design. In the experiment, we also included a 'no movement' (i.e. static) object and a 'novel' movement pairing in which the objects and movements were mismatched with a movement from one of the other learned objects. As

such, although the objects and motions were each separately familiar, their pairings were novel in this condition. Sounds, whether correlated or uncorrelated, were presented in all movement pairing conditions. Note that in the 'static' condition, although no movement was shown, the same (correlated and uncorrelated) sounds were presented with these images as in the object movement pairing condition. The ‘correlated’ sound in the static condition was the same as the learned sound pairing.

Prior to testing, each moving object was assigned to one of two categories with 3 objects per category, and category membership was counterbalanced across all participants. We also counterbalanced the labelling of the categories so that the pink planet (positioned on the left of each display) represented the ‘Bloop’ category for half of the participants and the ‘Neem’ category for the other half.

There were two separate sessions to the experiment: a category learning session followed by a categorisation test. In the category learning session, six moving objects were presented with correlated sounds and participants learned to categorise these objects. This task was based on one between-group factor of children’s 'age group' with three age groups of 5-8 years, 9-10 years, and 11-13 years. We also tested an adult group to compare performance to that of the children. During this session, we measured the number of trials required by each participant to reach a predefined learning criterion of 70% accuracy.

The categorisation test was based on a fully-factorial, mixed design with one between subjects’ factor and two within subjects’ factors. The between subjects’ factor was age-group (4 levels, including adults). The within subjects’ factors were object-movement pairings (3 levels: learned, novel and static) and movement sounds (2 levels: correlated or uncorrelated), yielding a total of 6 possible object conditions. All the six objects were presented in each of the six learned (movement by sound) conditions, yielding a total of 36

trials. In the categorisation task, the dependent variables were categorisation accuracy and reaction times.

Procedure

Each participant or their parent/guardian first provided informed consent (and assent in the case of children) before taking part in the study. On entering the test space, the participant was invited to sit at a table, and the laptop screen was positioned approximately 45cm away. First the participant was presented with written instructions on the screen which the experimenter also read aloud to the child. The task narrative for the children was to help each character to return to its home planet. There were two planets, named Neem and Bloop, each with their own colour and position on the screen. The task was to look at each object and decide which planet it belonged to by pressing the relevant ('a' or 'l') key on the keyboard that was closest to the planet (e.g. if the correct category was 'Bloop' and that planet was shown on the left, then the participant was instructed to press the 'a' key). Following these instructions, and to satisfy the experimenter that the participant understood the requirements of the task, they were initially presented with a trial in which they had to return a rocket ship to a named planet by pressing the 'a' or 'l' key. If this was successful, the experimenter then addressed any questions about the task that the participant raised, or they proceeded to the next step. This step involved an exposure phase in which each of the six moving objects were shown to the participants (in random order). The purpose of this was to allow the participant to see how many different objects and movement and sound types there were and to instruct the participant to attend to all characteristics of the objects during the experiment.

The experiment then proceeded to the category learning task in which participants were presented with the individual moving objects and their correlated sounds and were required to categorise each object by choosing the correct planet to which they belonged.

Both visual and auditory feedback was given on the accuracy of each response. Participants were required to complete at least 18 trials (i.e. 3 repeats of each of the 6 objects) and accuracy was calculated over a sliding window of the most recent 12 trials. The threshold for successful learning was 70% accuracy within this window. Additional trials were presented until the participant met the learning criterion. The maximum possible number of trials in the learning phase was capped at 96 trials (i.e. 16 repeats of each object) to allow the children to learn without being fatigued. If learning was not achieved by 96 trials, then the experiment was terminated for that participant. A self-timed break was offered after every 24 trials and for the child participants only, the experimenter remained present to maintain their focus on the task.

In the categorisation task, participants were asked to categorise the objects based on their visual shape. The objects were presented either with the same or mismatched movement relative to the learned movement, or they were presented without movement (i.e. as a static image). In turn, each object (moving or static) was also presented with a sound which was correlated or uncorrelated to their movement (apart from in the static condition in which the 'correlated' sound was the same as in the learning session). There was a total of 36 trials (6 objects shown once per each of the 6 movement-sound conditions) in the categorisation task.

Both the learning trials and test trials shared a similar structure. Each trial started with the presentation of a fixation cross for a duration of 250ms. Participants were instructed to view the fixation cross at the start of each trial. The object stimulus then appeared as either an animation (video clip with sounds) or static image (with sounds) and shown for a duration of 6 seconds. The participant could not respond while the stimulus was displayed. A visual reminder of the category labels appeared on the screen after the stimulus had ended and remained on screen until the participant responded. A response was

provided by pressing either the 'a' or 'l' key on the keyboard (indicated by coloured stickers). There was no limit on the time to respond although the experimenter prompted a response if there was a long delay. The experiment took approximately 20 minutes for each participant to complete.

Results

During the learning task, 12 participants (11 children and one adult) failed to reach the learning criterion threshold within the trial limit, leading to a final sample size of 128 participants. Of the child participants who did not learn, nine were 5-8 years (M = 7.76, SD = 0.94; three boys), two were 9-10 years (M = 9.53, SD = 0.04; one boy) and one was 11-13 years (12.3yrs, boy). Further details on participant numbers per age group are provided in Table 3.

Table 3

Details on age of the participant groups who reached the learning criterion

Age Group	N	*M (years)*	*SD (years)*	Min	Max	% Males
5-8 years	43	7.63	0.90	5.25	8.92	47
9-10 years	36	10.1	0.53	9.02	10.9	51
11-13 years	49	12.1	0.55	11.0	12.9	36
Adult	36	25.4	3.61	20.4	32.9	36

Category Learning performance

The mean number of trials needed to reach criterion for each of the age groups was as follows: 5–8-year-olds, M = 29.35, SD = 15.31 trials; 9–10-year-olds, M = 30.31, SD = 17.72 trials; 11–13-year-olds, M = 25.53, SD = 14.07 trials; and adults M = 30.85, SD = 12.46 trials. Overall, participants required approximately 5 repetitions of each object to learn the categories. In order to investigate the effects of age group on the number of trials (i.e. count data) required to reach the learning criterion we first conducted a Shapiro Wilk test which revealed the data were not normally distributed, W = 0.89, $p < 0.001$ and

Levene's test indicated a heterogeneity of variances [$F(3, 158) = 2.87, p = .038$]. A Kruskal–Wallis test was therefore used to compare performance across the age groups. The results indicated a non-significant effect of age group on the number of trials needed to reach criterion [$\chi^2(3) = 5.10, p = 0.16$].

Categorisation Test performance

Participants' data from the test session were excluded if they scored below chance overall or if they scored at chance to the trials in which the learned object-motion pairs with a correlated sound only were presented. These exclusions were considered necessary since chance performance to the learned object conditions suggested that the participant had either not learned the objects or were not attending to the task. This led to the further exclusion of data from 13 participants across all age groups: three 5–8-year-olds, six 9–10-year-olds, three 11–13-year-olds and one adult. These total exclusions represented less than 8% of the data from the participants who reached the learning criteria (see Supplemental Materials 3 for more details).

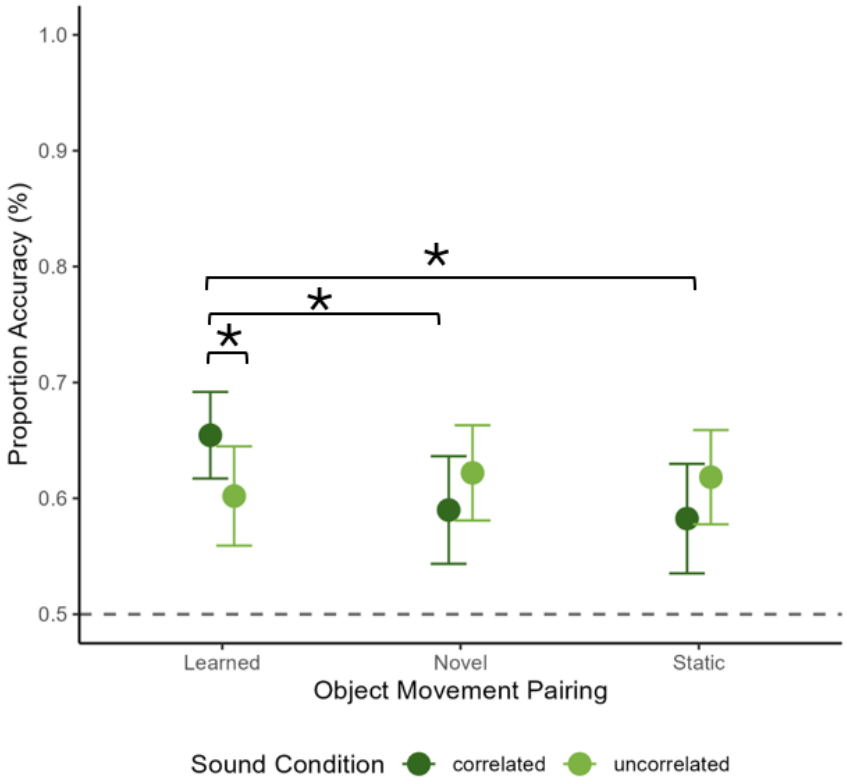
The children's accuracy and response times data were initially analysed separately to the data from the adults (a comparison between children and adult performance is described later). Categorisation accuracy performance was assessed using a mixed ANOVA with the between subjects' factor of age group (3 levels: 5–8 years; 9–10 years; 11–13 years) and within subjects factor of object-movement pairings (3 levels: learned, novel; static) and sound (2 levels: correlated; uncorrelated/unlearned). The dependent variable of interest in this initial analysis was accuracy (proportion). The assumption of sphericity was met for all within-subject effects, as indicated by Mauchly's test (all $p \geq .65$).

The main effects of age group [$F(2, 113) = 1.20, p = .305$, generalized $\eta^2 = .008$], object-movement pairing [$F(2, 226) = 1.18, p = .309$, generalized $\eta^2 = .003$], and sounds [$F(1, 113) = 0.11, p = .746$, generalized $\eta^2 < .001$] did not reach the level of statistical

significance. The two-way interactions between age group and movement pairing [$F(4, 226) = 0.49$, $p = .742$, generalized $\eta^2 = .002$], and between age group and sounds [$F(2, 113) = 0.47$, $p = .627$, generalized $\eta^2 = .001$] did not reach significance. However, there was a significant interaction between object-movement pairing and sound [$F(1.98, 224.29) = 3.84$, $p = .023$, generalized $\eta^2 = .008$], suggesting that the effect of movement pairing differed depending on whether the sound was correlated or uncorrelated (see Figure 8). Pairwise comparisons (Bonferroni-adjusted) revealed that when the sounds were correlated with the object movement, accuracy was significantly higher for learned object-movement pairings ($M = 0.65$, $SE = 0.02$) than for either the novel movement pairings [($M = 0.59$, $SE = 0.02$); $t(113) = 2.49$, $p = .043$] or the static objects [($M = 0.58$, $SE = 0.02$); $t(113) = 2.56$, $p = .035$]. No significant differences emerged between static and novel movement conditions. When sounds were uncorrelated, none of the pairwise comparisons between the movement conditions reached significance after Bonferroni correction (all $p \geq .193$). The three-way interaction among age group, movement pairing, and sounds was not significant [$F(4, 226) = 0.51$, $p = .727$, generalized $\eta^2 = .002$].

Figure 8

Plot showing the children's categorisation accuracy (proportion) across object-movement pairing and sound conditions. The accuracy data are averaged across age groups as age had no effect on performance.



Note. Error bars represent 95% Confidence Intervals. The dashed line represents chance level performance.

We decided to analyse response times although it is noted that these times were constrained by the duration of the stimulus presentation and the instructions. Nevertheless, we thought that response times might provide a meaningful insight into children's categorisation performance. The analysis of reaction times (RTs) was based only on trials to which the participant provided the correct response. We also removed any response times that exceeded ±3SD relative to each participant's mean reaction time, which led to the removal of 2.37% of data. As a consequence of this data cleaning procedure, the data from an additional six participants were not included in subsequent analyses as there was insufficient data remaining in one of their within subjects factor combinations for analysis.

Prior to analysis, the distribution of the mean reaction times (RTs) was examined to evaluate normality. Descriptive indices indicated a strong positive skew (skewness = 7.33) and leptokurtosis (kurtosis = 58.13). The Shapiro–Wilk test confirmed significant departure from normality [W = 0.15, p < .001]. Given these violations, a log transformation was applied to the children's RT data, consistent with standard practice for positively skewed reaction time distributions. All analyses reported below were conducted on the log-transformed RT variable.

Categorisation reaction time was assessed using a mixed ANOVA with the between subjects factor of age group (5–8 years; 9–10 years; 11–13 years) and within subjects factor of object-movement pairing (learned; novel; static) and sound type (correlated; uncorrelated). The dependent variable of interest was the mean log reaction time. Mauchly's test indicated that the assumption of sphericity was met for the main effect of object-movement pairing, $W = 0.96$, $p = .12$, and the age group by movement pairing interaction, $W = 0.96$, $p = .12$. However, sphericity was violated for the object-movement pairing by sound type interaction, $W = 0.94$, $p = .030$, and for the three-way interaction, $W = 0.94$, $p = .030$. Accordingly, Greenhouse–Geisser corrected degrees of freedom are reported where appropriate.

The effects of age group [$F(2, 107) = 0.48$, $p = .617$, generalized $\eta^2 = .007$] and sound type [$F(1, 107) = 0.02$, $p = .901$, generalized $\eta^2 < .001$] were not significant. There was a main effect of object-movement pairing [$F(1.90, 203.5) = 4.03$, $p = .021$, generalized $\eta^2 = .004$], indicating differences in RTs across static, learned and novel object-movement pairings. Post hoc pairwise comparisons with Bonferroni adjustment showed that children responded significantly faster to learned ($M = 0.09$ logRT, $SE = 0.16$) compared to novel ($M = 0.22$ logRT, $SE = 0.15$) object-movement pairings [$t(107) = –3.11$, $p = .007$]. No

significant differences were observed between static (M = 0.14 logRT, SE = 0.16) and learned movements (p = .429) or between static and novel movements (p = .619).

The pairwise interactions between age group and object-movement pairing [F(3.80, 203.48) = 0.93, p = .446, generalized η^2 = .002], age group and sounds [F(2, 107) = 0.14, p = .867, generalized η^2 < .001] and movement pairing and sound [F(1.83, 196.06) = 3.03, p = .055, generalized η^2 = .003] failed to reach significance. The three-way interaction between age group, movement pairing, and sound was also not significant [F(3.66, 196.6) = 0.73, p = .563, generalized η^2 < .001] for full details of mean and reaction times across all movement and sound conditions see Supplemental Materials 4.

Since we found no effect of age on children's performance, we decided to group all children together (N=116, mean age 10.1 years) to compare children's overall categorisation performance to that of adults (N=36, mean age 25.3 years). A linear mixed-effects model (LMM) was fitted to investigate the mean proportion accuracy across the factors of object-movement pairing (static; learned; novel) and sound type (correlated, uncorrelated), with age group (children or adults) as a between-subjects factor. In the first of the analyses, we used categorisation accuracy (mean proportion correct) as the dependent variable. Fixed effects included movement pairing, sound type, and age group. A random intercept for participant was specified to account for individual variance across repeated measures. Models were estimated using restricted maximum likelihood (REML), and significance testing of fixed effects was based on Satterthwaite's approximations of degrees of freedom.

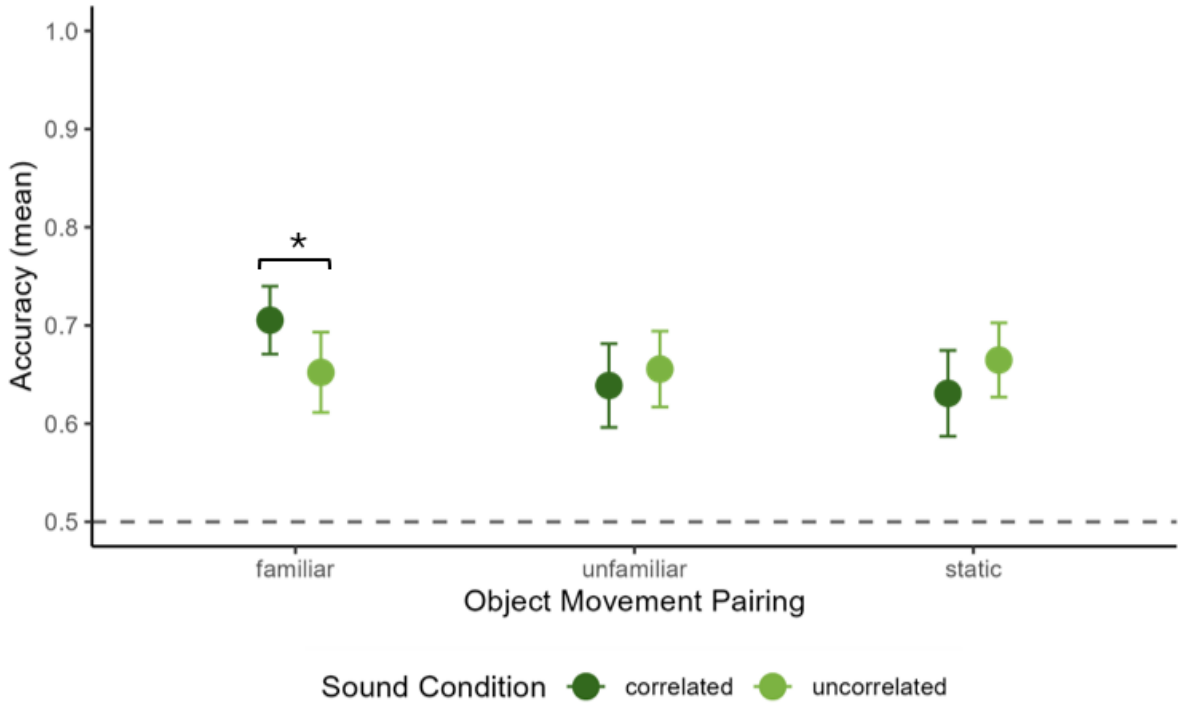
The model showed a conditional R^2 of 0.41, indicating that approximately 41% of the variance in accuracy was explained by both fixed and random effects, while the marginal R^2 was 0.12, indicating that fixed effects alone explained about 12% of the variance. Random effect variance for participants was estimated at 0.018 (SD = 0.14), with

a residual variance of 0.037 (SD = 0.19). Type III Wald χ^2 tests revealed significant main effects of object-movement pairing [$\chi^2(2) = 9.74$, $p = .008$, partial $\eta^2 = .007$]: accuracy was highest for learned (M = 0.74, 95% CI [0.70, 0.78]), than static (M = 0.70, 95% CI [0.66, 0.74]) or novel (M = 0.70, 95% CI [0.66, 0.73]) movements, but these differences failed to reach significance following Bonferroni correction for multiple comparisons. There was also a main effect of age group [$\chi^2(1) = 18.46$, $p < .001$, partial $\eta^2 = .21$] with more accurate categorisation performance by adults (M = 0.81, 95% CI [0.74, 0.89]) than children (M = 0.61, 95% CI [0.58, 0.64]). Children's proportional accuracy performance ranged from 0.59 to 0.66 across conditions, while adults ranged from 0.77 to 0.87. The main effect of sound was not significant, $\chi^2(1) = 1.92$, $p = .167$, partial $\eta^2 < .001$.

A significant object-movement pairing by sound interaction was observed [$\chi^2(2) = 7.34$, $p = .025$, partial $\eta^2 = .007$] on accuracy performance, as shown in Figure 9. Pairwise comparisons were then conducted for each level of sound type. For the correlated sound condition, accuracy for learned object-movement pairing (M = 0.66, 95% CI [0.62, 0.70]) was significantly higher than for static images [(M = 0.59, 95% CI [0.54, 0.63]; difference = 0.08, SE = 0.03; t(730) = 2.88, $p = .012$ (Bonferroni-adjusted)] and for novel object-movement pairings (difference = 0.07, p = .050). There was no statistical difference in accuracy between static and novel object-movement pairings (difference = −0.01, $p = 1.00$). A comparison of performance in the uncorrelated sound conditions suggested that none of the pairwise contrasts between object-movement pairings was significant (all ps ≥ .63). Finally, neither the object-movement pairings by age group interaction [$\chi^2(2) = 0.05$, $p = .97$, partial $\eta^2 < .001$], the sound by age group interaction [$\chi^2(1) = 0.01$, p = .95, partial $\eta^2 < .001$], nor the three-way interaction [$\chi^2(2) = 0.60$, p = .74, partial $\eta^2 < .001$] were significant.

Figure 9

Plot showing the interaction between sound and movement conditions on the mean proportion accuracy, derived from estimated marginal means (EMMs) collapsed across children and adults



Note. Error Bars represent 95% confidence intervals.

An analysis of children's and adult response times revealed the expected effects of age, with faster categorisation times by adults than children. None of the main effects (object-movement pairings or sound) had an effect, nor was there evidence of an interaction between the factors. Further details can be found under Supplemental Materials 5.

Discussion

The primary aim of this study was to investigate whether object motion and sound influence the learning of novel object categories and the subsequent categorisation of those objects across development. Specifically, we examined how children aged between 5 and 13 years of age, and adults, learn novel categories of moving objects that make a sound whilst moving. We then tested how these object representations, during acquisition of which a combination of motion and sound cues were present, influenced object categorisation when the movement of the object was familiar, unfamiliar, or no movement

is presented and when the sound was either correlated or uninformative (uncorrelated) with the object movement. We wished to assess whether the motion and sound cues independently support object categorisation when shape cues were held constant. It was also hypothesised that the influence of the motion and sound cues to categorisation may change with development, such that the performance of older children and adults would benefit more from correlated presentation of audiovisual cues during category learning and subsequent categorisation judgements than younger children.

Across the learning phase, there was no significant difference in the number of trials required to reach the learning criterion for successful category acquisition across all children's age groups and adults. This may not be too surprising given that the requirement of the learning task included a minimum exposure to 3 repetitions of each object. Most participants required between 5 and 6 repetitions before reaching the learning criterion. There was additionally no overall effect of children's age on accuracy during the categorisation test. This finding was in contrast to our initial hypotheses that categorisation would improve with age. Although the effect of age on categorisation with audiovisual object stimuli has not previously been tested (note that Broadbent et al. 2017 tested category learning), age-related differences have previously been reported for tasks involving visual object recognition (Ayzenberg & Behrmann, 2024; Bova et al., 2007; Dekker et al., 2011). In that context we expected an improvement in task performance with age. A possible interpretation of this absence of age-related differences is that due to the observation that as learning performance was similar across all age groups it may have supported equivalent categorisation performance across all children.

It is possible that the three child age groups (5–8, 9–10, and 11–13 years) fall within a relatively extended developmental period during which the functional use of multisensory information changes only gradually (Downing et al., 2015). Previous research suggests that

multisensory processing follows a protracted developmental trajectory, with adult-like behavioural benefits associated with integration, for example decreased reaction times emerging relatively late around age 14 (Brandwein et al., 2011). However, an alternative interpretation would be that performance in the present task may rely less on obligatory multisensory integration and more on the formation of cross-modal associative links between object shape, motion, and sound during learning. Developmental accounts emphasise that children must learn correspondences between sensory signals through experience, gradually building associations across modalities (Lewkowicz & Kraebel, 2004; Lewkowicz, 2014). Empirical work further shows that audiovisual associative learning can support category learning and memory even in the absence of optimal integration mechanisms (Broadbent, White et al., 2017). From this perspective, children across all age groups may be similarly capable of forming and retrieving these associations, resulting in comparable categorisation performance at test (Sloutsky, 2010). Furthermore, the characteristics of the stimuli utilised in this task may have reduced sensitivity to developmental differences. In particular, the use of multiple, temporally aligned and redundant cues (shape, motion, and sound) may have supported performance across all age groups. According to the Intersensory Redundancy Hypothesis, redundant multisensory information enhances attention to shared properties and facilitates learning from infancy (Bahrick & Lickliter, 2000; Bahrick, et al., 2004). At the same time, such conditions are also well suited to promoting associative binding across modalities (Lewkowicz & Kraebel, 2004; Lewkowicz, 2014), which may reduce developmental differences if children of different ages are equally sensitive to consistent co-occurrences of features. In addition, the use of a learning-to-criterion procedure ensured that all participants reached a comparable level of performance during category acquisition, which may have further reduced group differences at test. Previous research suggests that when task demands are reduced or

learning conditions support simpler associative strategies, children can perform comparably to adults (Minda et al., 2008).

To contextualise the observed results in our developmental group, we also measured categorisation performance in adults to provide a point of comparison. As expected, adults outperformed children in overall categorisation accuracy at test (Minda & Smith, 2011). As such, developmental differences were found but only between the children (as a group) and adults. We further predicted that both object-movement type and sound pairing would influence categorisation differently across the age groups. However, we found no evidence that age influenced performance in these conditions: the performance of both children and adult groups was similar to the object-motion and sounds.

However, we did find an interaction between object-movement pairing and sound (irrespective of age). Specifically, we found greater categorisation accuracy to objects presented under the same movement and sound conditions as learning relative to all other movement and sound conditions. In other words, the categorisation performance of both children and adults benefitted from the combination of object shape, object motion and the sound of the motion during the test and was significantly reduced if any one of the motion or sound cues differed from learning. Thus, this finding suggests that both movement and sound cues were perhaps associated into the category representations of the objects during learning. Importantly, however, the pattern of effects in the adults across the object movement–sound pairings mirrored that observed in the children. This may suggest that the categorisation processes based on object, motion and sound cues, were shared across children and adults, although overall adults were more efficient at the task. A potential interpretation of this observed difference between children and adults at categorisation test is a transition from an associative crossmodal representation in children, to an integrated multisensory representation of the newly acquired categories in adults. This interpretation,

in need of further investigation, could provide an interesting avenue for potential future research.

One possible interpretation of the finding that the learned cue combination led to better categorisation performance is that it reflects a familiarity effect, as this was the condition experienced during learning. However, this explanation alone could be considered to be unlikely. Firstly, when objects were paired with a movement that had previously been associated with a different object during learning, this movement feature was not novel to participants. For example, a participant may have learned that object 1 was rolling but then be presented with a different object rolling during the categorisation test: in this example, a rolling movement is familiar in both contexts. If familiarity alone were driving performance, we would expect similar performance across the object-motion conditions, which was not observed. Secondly, the presence of a correlated sound did not enhance performance, even in our youngest participants who should be capable of discriminating sounds (Robinson & Sloutsky, 2004a). Indeed, the presences of a correlated or non-informative sound had no effect on the categorisation of novel object-movement pairings. A relative benefit only emerged when the learned movement was paired with its correlated sound, indicating that it was the specific audiovisual pairing learned during training, and not familiarity with individual features, which supported categorisation. This may suggest that both children and adults associated the object–movement–sound cues into the category representations which facilitated subsequent categorisation.

These findings align with prior work demonstrating that object motion is an important cue for object identification and discrimination, and that movement information supports object representations in memory (Chung et al., 2022). Moreover, previous evidence suggests that motion can facilitate recognition when static information is ambiguous or degraded (Pavlova et al., 2001), and that sounds associated with an object's

movement can also aid perceptual judgements (Li & Deng, 2023; Chen & Spence, 2010). Importantly, our findings extend this work by suggesting that motion cues are also associated into memory representations that support later shape-based categorisation. Finally, the finding that unfamiliar motion and uncorrelated sound reduced categorisation performance, particularly in children, highlights the potential importance of cross-modal cue association in supporting category learning and recognition.

It is important to note that performance was above chance for all conditions, even when the movement or the sound differed from the learned version, suggesting that both children and adults could base their categorisation on the object shape alone. Moreover, object motion and sound were unreliable cues to categorisation during the test, because these features were changed across the different objects, and object shape was the only reliable cue for categorisation. Therefore, shape appeared to be successfully selected by all children for the categorisation task. Nevertheless, both children and adults seemed to attend to the motion and sound features since the learned combination of these cues benefited performance. This result suggests that both children and adults could adapt their category representations to successfully generalise to novel exemplars.

Although we found that correlated sound benefits categorisation, the specific role of the sound is unclear. For example, our findings cannot disambiguate between the role of sound as an independent cue to categorisation or whether sound increased the saliency of the motion itself during learning. One way to address this might be to test the effect of sound when object categories are acquired with uncorrelated sounds or without sound and then introduce a correlated sound at the test stage. Alternatively, learning efficacy could be assessed to shape alone compared to the combined cue conditions of shape and movement or shape, movement, and sound. Moreover, it would be of interest to investigate the impact of multiple training sessions across different ages to assess learning rates as well as simply

number of trials to reach criterion. This may allow us to characterise potential transitional changes to strategy use as well as underlying cognitive capacity.

In conclusion, the present study provides novel evidence to support that both children and adults may associate audiovisual motion information and shape cues during category learning and that these audiovisual associations influence subsequent categorisation performance. While children and adults achieved similar learning outcomes in terms of trials to criterion, their reliance on sensory cues differed, possibly reflecting changes in how multiple cues are stored in memory across development. The finding that categorisation benefits from a familiar combination of object movement and sound may suggest some specificity of these representations; nonetheless, the generalisation performance observed to novel combinations of object movement and sounds suggests that categorisation can be supported when shape cues are constant. These results contribute to our understanding of how crossmodal information influences category learning in children and adults.

Chapter 4

Can children generalise their acquired object categories to novel exemplars defined by visual, haptic and visuohaptic features?

Chapter 4

Abstract

Although children constantly interact with the objects in their environment using vision and touch, our understanding of how these senses contribute to category formation is relatively poor. Children may find the complementary information about the relevant object features across vision and touch helpful for categorisation. Alternatively, cross-sensory information may increase the search space within which the relevant category features are encoded. The aim of the current experiment was to investigate children's ability to learn and generalise object categories based on unimodal or crossmodal information. Different groups of children aged from 4 to 13 years (N = 245) first learned to categories novel objects using either vision (V), haptics (H) or both (VH). Category learning was poor in younger children particularly to the VH modality, but modality had no effect on learning in older children. Categorisation performance was then assessed using learned and novel exemplars defined by a change of object features selected from either within or across categories. Generalisation performance was unaffected by within-category feature changes, but was reduced with cross-category feature changes, particularly those defined by haptic information. Our results suggest that children's ability to generalise object categories is best for visual information. Contrary to our predictions, we found no specific benefit for multisensory information on category learning or generalisation at any age. These findings provide insights into children's ability to selectively attend and encode object features and suggest an independent influence of vision and touch in the formation of object categories.

Chapter 4

Introduction

In their daily lives, children are constantly interacting with objects that typically comprise featural information which is accessible by sight and touch. The organisation of this multisensory featural information into object categories is a process which is critical for learning, the generalisation of previous experience to novel situations, and the development of vital cognitive skills. Although there has been significant progress on our understanding of children's ability to form object categories based on visual information alone (see e.g. Mareschal & Quinn, 2001; Spriet et al., 2022; Ayzenberg & Behrmann, 2024; Quinn et al., 1993), it is important to account for the contribution of object information from other sensory systems on this everyday ability. This is especially true for touch which is an expert system for identifying objects (Klatzky, Lederman & Metzger, 1985) as well as being critical for learning in young children (Denner and Cashdan, 1967). Children's ability to effectively utilize featural information from the visual and tactile senses in the formation of novel categories and subsequent generalisation is, however, poorly understood.

Each of the perceptual steps involved in the formation of object categories is subject to developmental effects. To form novel categories, the child must first selectively attend and encode relevant object features (Shepard et al., 1961). Studies of selective attention suggests that this process becomes more efficient with age although children younger than 6 years of age rely more on distributed attention to make object judgements (Sloutsky & Deng, 2017). The process of categorising objects for the child is further complicated by the fact that not all object features are diagnostic of category membership. For example, although a 'tail' feature is common to many animals this feature alone would be insufficient to distinguish cats from dogs. Furthermore, feature selection must be adaptive to the task (whilst a 'tail' feature may not help some category tasks it may be useful for others, such as categorising pigs from dogs). Thus, both prior knowledge and adaptive encoding are

important for relevant feature selection. Object categories are then formed by integrating these features into category representations that maximise differences across, whilst minimising differences within, categories (Deng et al. 2016; Goldstone, 1994; Nosofsky, 1986). Although young children can already form categories on the basis of similarity (Sloutsky, 2003) other types of categorisation, such as rule instantiation, rely on the development of cognitive processes (Rabi & Minda, 2014).

Object features encoded through each modality can be either complementary or correlational and each type can influence categorisation (Newell et al., 2024). For example, some object features may be accessible only to vision or to touch, such as object colour or weight respectively, whereas other features may be accessible to both senses, and therefore redundant across modalities, such as surface texture and shape. Infants as young as just a few days old appear to demonstrate stimulus equivalence across vision and touch for object features such as shape and texture (Streri et al., 2000; Streri & Gentaz, 2004; Molina & Jouen, 1998), suggesting that categorisation may benefit from featural redundancy across childhood.

Indeed, some findings suggest a benefit for intersensory redundancy on attention (e.g. Bahrick, Lickliter & Flom, 2004) leading to improved perception for multisensory versus unisensory information even in infants (Bahrick and Lickliter, 2000; Baker & Jordan, 2015; Jordan & Baker, 2011). However, up until recently, studies exploring the facilitative effects of multisensory exposure on category learning were mostly conducted in the audiovisual domain. For example, Broadbent et al. (2017) reported that children older than 6 years demonstrated a multisensory benefit during an incidental, audiovisual category learning paradigm and this benefit on learning increased with age. Moreover, a benefit on categorisation following multisensory learning persisted over a delayed retention of 24 hours (Broadbent et al., 2019). In addition, Kirkham et al. (2019) reported a benefit on

category learning with audiovisual features in children from the age of 5 years old, although visual cues were better retained as category representations in memory in both 5- and 10-year-olds.

Featural information encoded through multiple modalities can, however, increase the search space from which the child can select the relevant features. To mitigate against this complexity, there is some suggestion of sensory dominance or bias across vision and touch in young children especially (e.g. Millar, 1971). For example, perceptual judgements of object size have been found to be haptic dominant, whereas vision is the more dominant sense for perception of orientation (Gori et al., 2008; 2010), in accordance with relative sensory precision. Moreover, the perception of object shape through touch is dependent on development: from the age of 2.5 years children can recognise objects using touch alone (Bigelow, 1981; Bushnell et al., 1999) and performance improves with age as a consequence of more efficient haptic exploration of object features (Morrongiello et al., 1994). Consequently, the influence that each sensory system has on object categorisation is likely to change with development in accordance with its rate of maturation and the precision of the encoded featural information (Misceo et al., 1999).

In a recent study, Broadbent et al. (2020a) reported that visuohaptic exposure benefitted object category learning of novel objects over either modality alone. However, this benefit was found in older children from the age of 8 years, the age at which children can optimally integrate information across visual and haptic modalities (Gori et al. 2008). Consistent with previous evidence for sensory dominance in young children (Nava & Pavani, 2013; Lewkowicz, 1988b; Gori, Giuliana et al. 2012; Robinson & Sloutsky, 2004a), Broadbent et al. also reported a greater benefit for haptic-only than visual-only featural information in young children (6-7 years), with visual learning alone leading to poorer learning outcomes. The study did not, however, address whether these modality effects on

category learning subsequently influenced categorisation performance or generalisation to novel exemplars. The question remains of how exposure to distinct and redundant haptic and visual cues during category learning affects categorisation and is the focus of the current study.

Research Aims and Hypotheses

In the following study we adapted the earlier study reported by Broadbent et al. (2020a) to investigate children's ability to utilise visual and haptic featural information in the formation of novel object categories and subsequent generalisation. We tested children in three groups aged between 4 to 6, 7 to 9, and 10 to 13 years. The selected age groups (4–6, 7–9, and 10–13 years) were chosen to reflect potential periods of developmental transitions in multisensory cue use. Children under approximately 7–8 years typically rely on modality-specific dominance rather than integrating sensory information, consistent with accounts of crossmodal calibration (Gori et al. 2008; Gori, 2015), whereas children begin to show more reliable, though still developing, multisensory integration from around 8 years, with our middle age-group designed to capture a potential period of transition (Murray et al, 2016). Thus, these groupings are intended to capture a potential progression from modality dominance to increasingly adult-like multisensory processing across childhood. Children first learned to categorise novel objects in which some visual, haptic and visuohaptic features were informative of the object's category membership whereas other features were non-informative (i.e. distractors). Specifically, children were assigned to one of three learning modality conditions: a visual-only, haptic only, or visuohaptic condition. Broadbent et al. (2020a) found that children above 8-years-old showed enhanced category learning in multisensory compared to unisensory conditions, whereas younger children benefitted more from unimodal, specifically haptic, information. The goal of the first part of our study was to replicate this learning effect in that we expected older

(i.e. children aged 10 years old and older in our study) but not younger (i.e. children aged between 4 and 6-year-olds) children to have relatively better category learning performance (i.e. fewer trials to reach a learning criterion) in the visuohaptic compared to either unimodal condition.

We then assessed children's ability to generalise their learned unimodal or visuohaptic categories in a visuohaptic categorisation task. The aim was to compare children's ability to categorise learned objects, as well as to flexibly generalise their learning to categorise novel object exemplars as a consequence of the visual, haptic or bimodal learning modality. To investigate children's generalisation, the novel exemplars were designed such that one of the object's visuohaptic, haptic or visual features was changed with either a different within-category or cross-category feature. Thus, participants were asked to categorise previously learned objects in which all features were the same as in the learned objects, novel object exemplars defined by a different within-category feature or a different cross-category feature relative to the learned objects. This manipulation allowed us to investigate the nature of children's newly formed object categories and whether the modality of the relevant object features affected object generalisation. Moreover, we examined whether young children were more likely to find categorisation in the bimodal learning condition more difficult due to the larger feature space or if they would rely more on featural information from one modality, due to sensory dominance, as suggested by previous reports. Furthermore, we were interested in determining at what age category formation and generalisation benefits from the availability of multisensory object information.

Methods

Participants

We recruited 245 children aged between 4 and 13 years old to participate in the experiment from various educational institutions in the UK and Ireland. The final sample size was determined by an *a priori* power analysis (conducted using G*power 3.1) which was based on data from a similar study (Broadbent et al., 2020a). This calculation suggested that a minimum sample size of 234 participants would be necessary for an interaction with a medium effect size (*f*) of 0.25, $a = 0.05$, and power (1- β) = 0.80. A total of 36 children were recruited during the 'Summer Scientist' week hosted by the University of Nottingham, UK. The remaining 209 participants were recruited from 12 different primary schools from the greater Dublin area and surrounding counties in Ireland.

For the purpose of our experiment, the children were grouped into one of three age groups: 4 to 6-year-olds ($n = 85$ (38 boys); mean age = 5.91 years, age range = 4.08-6.97 years); 7 to 9-year-olds ($N = 76$ (42 boys); mean age = 8.39 years, age range = 7.00-9.92 years); and 10 to 13 year olds ($N = 84$, (45 boys); mean age = 10.95 years, age range = 10.00-13.8 years). Within each age group, the children were randomly assigned to one of three learning groups (see Table 4 for details).

Ethical approval for the experiment was given by the School of Psychology Research Ethics Committee at Trinity College Dublin. Accordingly, written consent was acquired from each child's parent or guardian, and each child assented to participating prior to the experiment. All participants' parents or guardians reported the normal or corrected-to-normal vision and normal hearing of their child.

Table 4

The number of participants in each age group allocated to each learning modality

Learning Modality	Age Group			
	4-6 years	7-9 years	10-13 years	**Total**
Visual	31	31	31	93

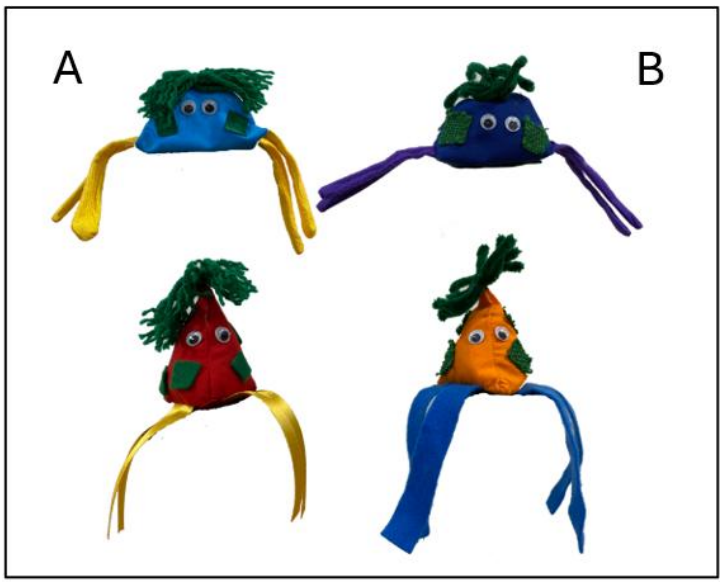
Haptic	25	23	30	78
Visuohaptic	29	22	23	74
Total	85	76	84	245

Stimuli and apparatus

We adapted the stimuli described by Broadbent et al. (2020a) to create a set of 54 individual novel objects ('alien creatures') for use as stimuli in our experiment. The objects were made from different types and colours of fabrics and were similarly structured in their shape (see Figure 10 for an example of 4 object stimuli). All objects comprised of a large body part, to which 'hair', 'legs' and body 'patches' were attached. In addition, we attached a pair of eyes to each object to make them more 'toy-like' and encourage the children to engage with the stimuli. The size of the objects was standardised so that their maximum dimensions were 10cm height x 10cm width x 5cm depth. As such they could be easily handled by child participants across the different ages**.**

Figure 10

An example of four (from the total of 54) object stimuli used in the experiment



Note. All objects were made of fabric and included features that were diagnostic of category membership: bodies, hair, legs, and patches. Category A objects ("Neems") are displayed on the left of the image and Category B ("Dorps") are displayed on the right of the image.

We manipulated the object features to create two different categories of objects. Specifically, the dimensions of four main object features were diagnostic of category membership. Of these four features, two were bimodal and therefore accessible using either vision or touch (i.e. fluffy or spiky hair type; rough or smooth textured patches), one was accessible to vision-only (i.e. leg colour), and one to touch only (i.e. light or heavy body weight). Each of the two object categories were defined by the following features, and there were two different versions of each feature per category:

Visual-only feature: The colour of the objects' legs was either warm (orange, yellow) or cool (blue, purple). The same colour was used for all the legs of any one object.

Haptic-only feature: An object's body weight was either light or heavy. Different materials were used for the light objects (cotton, plastic beads) and heavy objects (rice, chickpeas). The weight of the objects ranged from 6 to 90 grams: the light objects weighed and average of 8g and the heavy objects weighed an average of 78g.

Visuo-haptic feature 1: The material used for the objects' hair was either fluffy or spiky. Different strands of wool were used for 'fluffy' hair (fine, thick) and different materials were used for 'spiky' hair (wire lined felt, pipe cleaners).These fluffy/spiky dimensions were easily discriminable to vision or touch.

Visuo-haptic feature 2: Each object had seven 'patches' i.e. fabric rectangles measuring 1cmx3cm, which were equally distributed across the surface of the body. Different textures were used for the material of the patches which were either smooth (ribbon, felt) or rough (plastic mesh, hessian).

The objects also comprised four other features that were also defined by the modality which provided access to the information, but their dimensions were non-diagnostic of category membership. These included body colour (vision only), body compressibility (i.e. hard or soft, haptics only), leg texture and body shape (both visuo-

haptic). See Table 5 for a summary of how the features were allocated to different categories. All other incidental dimensions of a feature remained constant throughout the experiment. For example, for hair type, only the texture of the hair (spiky or fluffy) differed across categories, but hair colour remained constant. The non-diagnostic features were evenly distributed across the objects and counterbalanced across categories. Table S4 under Supplemental Materials 6 includes further details of the specific features used.

Table 5

A list of individual object features which were either diagnostic or non-diagnostic of category membership. Each feature type was accessible to either vision-only, haptics-only or to both vision and haptics (visuohaptic)

Feature modality	***Category Feature type***			
	Diagnostic			***Non-diagnostic***
Visual only	Leg colour	Warm *Orange/yellow*	Cold *Blue/purple*	Body colour
Haptics only	Weight	Light *Cotton/plastic*	Heavy *Rice/chickpea*	Compressibility
Visuo-haptic (1)	Hair	Fluffy *Fine/thick*	Spiky *Wire/pipe*	Leg texture
Visuo-haptic (2)	Patch	Smooth *Ribbon/felt*	Rough *Mesh/hessian*	Body shape

Of the total object stimulus set (54 objects), half of the objects belonged to Category A and the other to Category B, counterbalanced across the children. From the total set, 12 objects were assigned to the category learning phase, with 6 exemplar objects in each category. Each of these 12 objects included two different versions of each of the 4 main feature types that were diagnostic of category membership (e.g. either fine or thick 'fluffy' hair). The categorisation test also included the remaining 42 objects. These 42 objects were designed such that they each shared 3 of the 4 original diagnostic features as the learned counterpart, and this allowed us to manipulate the nature of the fourth feature. Specifically, for 24 of these 42 novel objects the fourth feature was changed and replaced by a different version of the original whilst still conforming to *within-category* feature dimensions. For example, if the original, learned object had 'spiky' hair then the new exemplar would also have 'spiky' hair but in a different material, or likewise, if the original object had a bright leg colour, then the novel feature would also be bright but in a different colour (e.g. yellow legs instead of orange). For the remaining 18 objects used in the test, the fourth feature was also changed but replaced by a related feature from the opposite category, i.e. a *cross-category* feature. In this case, for any one object, 3 of its diagnostic object features were

consistent with one category and 1 of the features was consistent with the other category. For example, if the original learned object had spiky hair (Category A), then the hair feature in the novel exemplar was fluffy (Category B) as shown in Table 2. Furthermore, across these 18 objects there were three different types of cross-category feature changes, each representing one of the learning modalities: visual-only (5 objects), haptic-only (5 objects) or visuo-haptic (8 objects) feature changes. See Table S1 under Supplementary Materials for details on the specific feature changes that were made to the objects in these conditions. This 'feature change' manipulation allowed is to assess children's ability to generalise to objects defined by novel feature configurations, and whether the encoding modality affected this generalisability.

The experimental apparatus consisted of a desk space on which two boxes were placed, one to the left and one to the right of the participant. Each box represented a specific ‘spaceship’ which was clearly indicated by a unique image of a spaceship (one grey and one black) fixed to each box facing the participant (see Figure 11). The allocation of the labels “Neem” and “Dorp” to the object categories as well as the ‘spaceship’ image used to represent each category was counterbalanced across participants. The objects were stored out of sight of participants, with the experimenter revealing each object, one by one, during the learning and testing trials.

In the haptic learning condition, an occlusion box was used which was positioned in the centre of the table and had two curtained openings in each side to allow the participant to place their hands inside the box. The rear of the box was open to allow the experimenter to place each object in the box for the participant to explore with their hands. All testing was conducted in a quiet room or area at each of the testing sites.

Figure 11

A schematic illustration of the experimental set-up

Note. The participant is seated on the left of the image and the experimenter (right) is presenting the object in the visual-only condition. Two boxes (the 'spaceships') were used to represent each of the object categories and the child's task was to indicate, once they explored and/or viewed the object, into which of the two boxes the object belonged.

Design

The experiment consisted of two tasks: a Category Learning phase followed by a Categorisation Test. Prior to the experiment, each participant was randomly allocated to one of three learning modality conditions for the Category Learning phase: visual only (V); haptic only (H); or visuohaptic (VH). These groupings resulted in a minimum of 21 participants allocated to each learning modality condition per age group. In contrast to learning, all children (i.e. across the different age groups) were presented with the same Categorisation Test in which the objects were explored using both vision and haptics.

In the Category Learning task, the experiment was based on a fully factorial, between-subject design with age group and learning modality as the factors. The age group factor included three age levels (4-6 years; 7-9 years; 10-13 years) and the learning

modality factor also included three levels (visual-only, haptic-only and visuohaptic learning). The trials were blocked in the Category Learning part of the experiment, such that all 12 objects (6 per category) were presented in one block of 12 trials (randomly ordered), and these blocks were repeated a maximum of 3 times to ensure learning. The outcome measures were the accuracy at categorising the 12 objects as well as the number of learning trials needed to reach a pre-defined learning criterion of 8 correct responses in a row (similar to Broadbent et al., 2020). This number of correct responses in a row corresponded to an overall categorisation accuracy of approximately 67% correct which is considered a reasonable performance criterion in category learning tasks for children (see e.g. Huang-Pollock et al., 2011).

In the Categorisation Test phase of the experiment the children were tested on their ability to categorise learned objects and generalise to novel object exemplars following visuohaptic object exploration. Only the children who had successfully learned to categorise the objects in the Category Learning phase (i.e. reached the learning criteria) participated in the categorisation task. The outcome measure for the Categorisation task was accuracy performance. In this task the children were presented with objects exemplars in which none of the object features were changed (6), one feature was changed but selected from within-category features (24), or one feature was changed but selected from cross-category (18) features across trials. The total of 48 objects in this task were presented in trials in a randomised order across participants. This 'feature change' condition was the main within-subject factor (none, within-category and cross-category) in this task. In addition, the cross-category feature change condition was further defined as a change to a visual-only, haptic-only and visuo-haptic feature relative to original learned features.

Procedure

The experiment was introduced as a tabletop game to the participants, using age-appropriate instructions which were verbally presented from a script to the participant (see Supplemental Materials 7 for details).

In the Category Learning task and depending on the learning modality condition the participant was assigned to, they were presented with each object either visually, using haptics only (out of sight) or they could both see and feel each object at the same time. The objects were presented for six seconds and then removed from view/their hands. In the learning phase, the participant was asked to respond as soon as the object was removed. In the visual condition the experimenter held the relevant object stimulus and extended it to the centre of the table, rotated it 180° to the left and then to the right and back to centre (see Figure 2) for six seconds. In the haptic condition, the child placed their hands within an occlusion box which was present in the centre of the table. The object was placed in the box next to the participant's hands and they were prompted to explore the object using both hands. After six seconds the child released the object from their hands, and the experimenter removed it from the box. In the visuohaptic condition, each object was handed to the child who then explored it visually and haptically for six seconds, after which the object was taken back by the experimenter and removed from view. For each trial, the child responded either by verbally indicating or by gesturing (pointing) towards the relevant spaceship. For all trials, the children were encouraged to respond immediately following the six seconds of object exploration, consistent with previous research on object categorisation in children and adults (Bushnell & Baxt, 1999; Gaissert & Wallraven, 2012). There was no limit on the time to respond but after an additional four seconds had lapsed, a verbal prompt to respond was given by the experimenter, if necessary. Both verbal and demonstrative feedback were provided after each learning trial (e.g., saying "Correct, that was a Neem" or "Incorrect, that time it was a Neem" and then gesturing towards the correct 'spaceship' or box).

If the participant reached the learning criterion, they immediately proceeded to the Categorisation Task. Before this task they were advised to remember what they had learned about the “Neem” and “Dorp” objects, but that sometimes a new object may appear which differs from those they had learned. They were also asked to recall the features that made an object ‘belong’ to either the Neem or Dorp spaceships (i.e. the features that made the Dorps or Neems “special”). We deemed this necessary to encourage participants not to rely on one feature alone, which may change or be absent in the categorisation task and encourage the recollection of other features if only one was mentioned by the child. In the Categorisation task the participants were presented with a set of objects across trials which included some that were originally learned in the learning phase, or novel objects in which one feature differed from the learned versions. Trials were presented in random order across participants and categorisation ability was tested using visuohaptic exploration only.

Whether or not a participant reached the learning criterion, the same debriefing procedures were conducted. The child was thanked for their time and offered a small token for participating. They were again asked what features they used to tell the two alien types apart, and the experimenter recorded their responses (see S5 for qualitative review of responses). Testing sessions lasted no more than 20 minutes per child.

Results

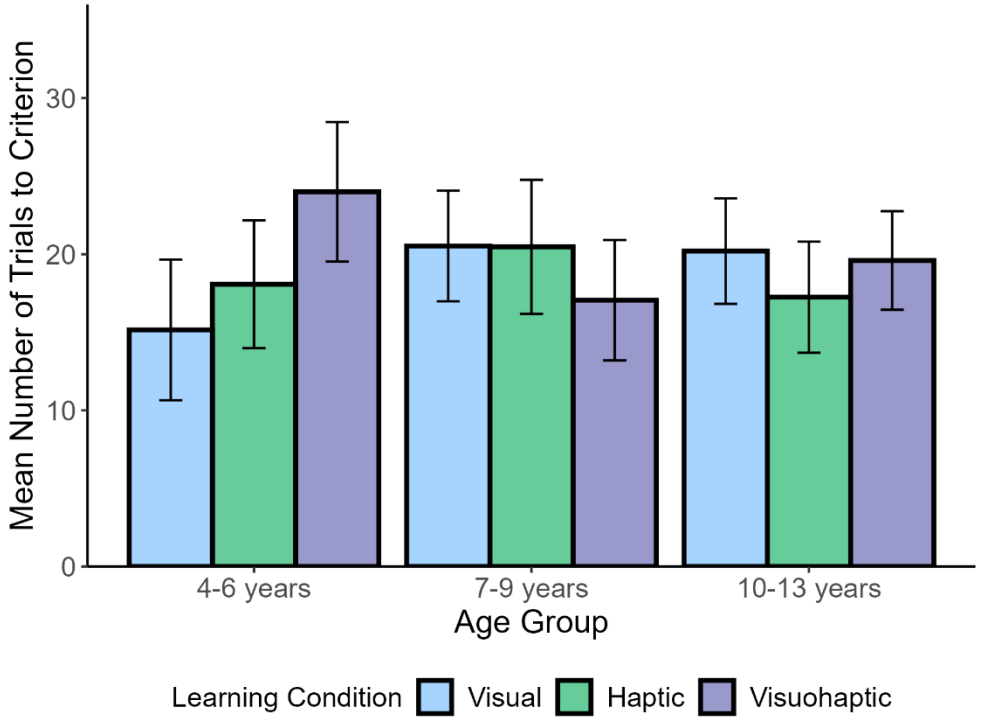
Category Learning performance

The number of children who reached the learning criteria in the Category Learning phase differed by age group. For the 4–6-year-olds, 41 reached the criteria but 44 failed to reach it; 60 of the 7–9-year-olds passed but 16 failed to reach the criterion and 66 of the 10–13-year-olds passed and 18 failed to reach the criterion. (Further analysis of the probability of passing the learning task can be found under Supplemental Materials 8). Of the children who passed the learning phase, the average number of trials needed for the

children to reach learning criterion overall 19.16 trials. See Figure 12 for the average number of trials needed to reach criterion by children across the different age groups in each of the learning modality conditions.

Figure 12

Plot showing the number of trials needed to reach the learning criterion for children in each of the age groups across the different learning modalities. Learning performance is shown only for children who successfully reached learning performance



To assess the efficacy of learning in the Category Learning phase, the number of trials needed to reach the learning criterion was compared across children in each of the age groups and learning modality conditions. A Shapiro-Wilk test was conducted which revealed that the distribution of the number of trials to criterion significantly deviated from normal, $W = 0.95$, $p < 0.001$. Therefore, an aligned rank transform ANOVA was conducted using the ARTool R package (Kay & Wobbrock, 2021). This method of analysis was chosen as it allows for the assessment of potential interactions between factors on non-continuous count data which is not normally distributed (Elkin et al., 2021). We examined the potential effects of learning modality (haptic, visual, visuohaptic) and age group (4-6 years; 7-9

years, 10-13 years) on the dependent variable of number of trials required to meet our learning criterion (8 correct responses in a row). The effects of learning modality [$F(2,156) = 0.17$, $p = .84$, $\eta^2_p = .003$] and age group [$F(2,156) = 0.01$, $p = .99$, $\eta^2_p < .001$], were not significant. However, there was a significant interaction between learning modality and age group [$F(4,156) = 2.96$, $p = .022$, $\eta^2_p = .07$] as shown in Figure 13. To explore this interaction, estimated marginal means and pairwise comparisons were conducted. Within the youngest age group (4–6 years), learning performance in the Visuohaptic condition suggested more trials were needed to reach criterion compared to the Visual-only condition, but this difference failed to reach statistical significance (difference = 10.55, $p = .059$, Bonferroni adjusted). No significant differences across modalities were found in two the older age groups (7–9 and 10–13 years; *ps* > .18).

Overall performance at Categorisation Test

The Categorisation Test phase included fewer children than the learning phase for several reasons. First, of the 245 children who originally participated in the Category Learning phase only 167 reached the category learning criterion (described above). An additional 7 participants from the younger age group (4–6-year-olds) dropped out of the study due to fatigue. Thus, a final sample size of 160 participants took part in the categorisation task. The number of these children in each of the age groups and learning modality conditions is shown in Table 6.

Table 6

Details of the participants who passed learning criterion and took part in the Categorisation Task

Learning modality	Age Group (in years)			
	4-6 (*n*)	7-9 (*n*)	10-13 (*n*)	Total (*N*)
Visual	11	19	20	50
Haptic	13	19	25	57
Visuohaptic	10	22	21	53
Total (N)	34	60	66	160

We first conducted a 3x3x3 mixed ANOVA with the between subjects' factors of age group (4–6 years, 7–9 years, 10–13 years) and learning modality (Haptic, Visual, Visuohaptic) and the within subjects' factor of feature change (none, within-category, cross-category). Mauchly's test indicated that the assumption of sphericity was violated for the feature change factor, [$\chi^2(2) = 16.23$, $p < .001$], and the associated interactions involving stimulus type, therefore Greenhouse-Geisser corrections were applied where appropriate.

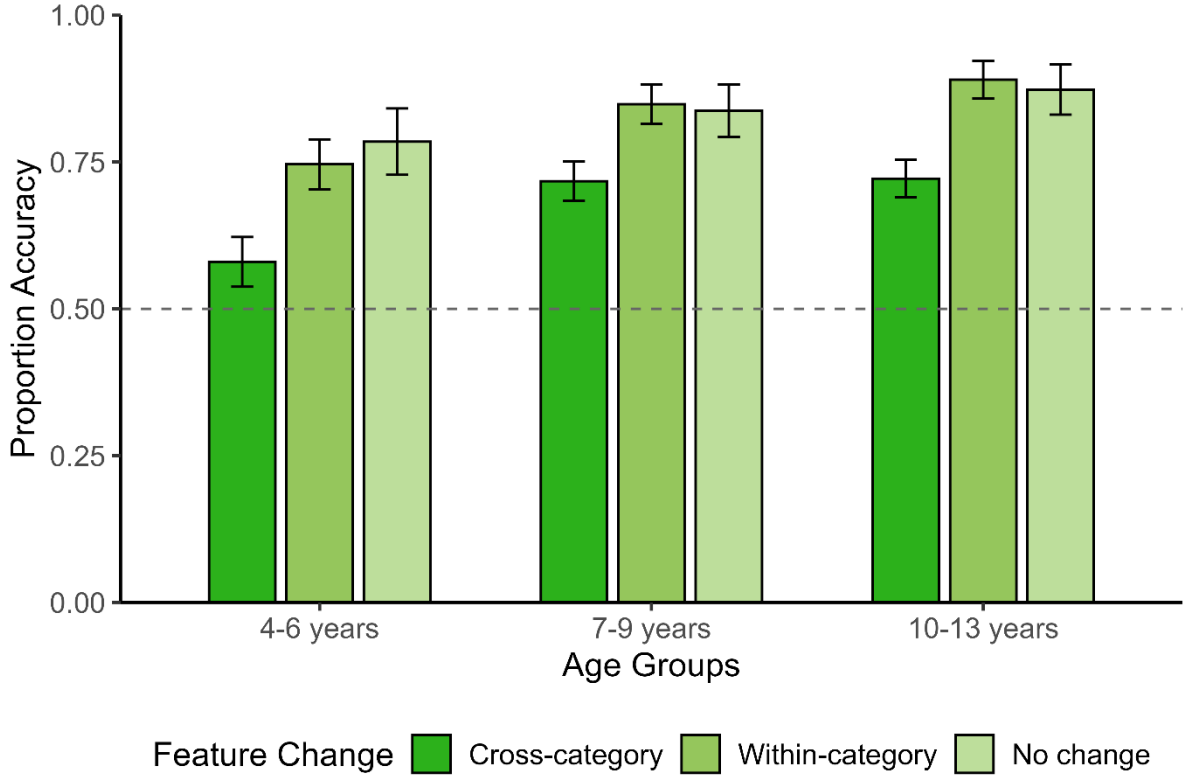
We found a main effect of age group [$F(2, 149) = 15.66$, $p < .001$, $\eta^2 = 0.11$] on categorisation accuracy. Pairwise comparisons using estimated marginal means showed that the youngest group (4–6 years; M = 0.70, SD = 0.16, 95% CI [0.67, 0.74]) performed significantly worse than both the 7–9 years-old group (M = 0.80, SD = 0.12, 95% CI [0.77, 0.83]; t(149) = -4.23, p = .0001) and the 10–13 years-old group (M = 0.83, SD = 0.07, 95% CI [0.80, 0.86]; t(149) = -5.51, p < .0001). There was no significant difference between the performance of the 7–9 and 10–13 years-old groups [t(149) = -1.39, p = .35].

The effect of learning modality was not significant [$F(2, 149) = 0.41$, $p = .66$, $\eta^2 = 0.003$], indicating comparable performance across Haptic, Visual, and Visuohaptic conditions which is an expected consequence of our learning criteria. There was a main effect of feature change [$F(2, 298) = 89.08$, $p < .001$, $\eta^2G = 0.21$] shown in Figure 13. Accuracy was highest for objects with either no feature change (M = 0.83, SD = 0.20, 95% CI [0.80, 0.86]) or a within-category feature change (M = 0.83, SD = 0.14, 95% CI [0.81, 0.85]), which did not differ significantly from each other [t(149) = -0.24, p = .97]. Performance was lowest for cross-category feature changes (M = 0.67, SD = 0.15, 95% CI [0.65, 0.69]) and significantly below both that of the no change [t(149) = -10.63, p < .0001] and within category feature changes [t(149) = -13.80, p < .0001].

None of the following pairwise interactions were significant: age group and feature change condition [$F(4, 298) = 1.75$, $p = .14$, $\eta^2G = 0.01$], age group and learning modality [$F(4, 149) = 1.19$, $p = .32$, $\eta^2G = 0.02$], learning modality and feature change [$F(4, 298) = 0.50$, $p = .74$, $\eta^2G = 0.003$]. The three-way interaction also failed to reach significance, [$F(8, 298) = 0.82$, $p = .59$, $\eta^2G = 0.009$].

Figure 13

Plot showing mean accuracy (proportion) in the categorisation test across type of feature change and age groups



Note. Error bars represent 95% CIs and the dashed line represents chance performance.

Generalisation across modality of cross-category feature change

The results of the analysis of performance across all categorisation trials suggested that a within-category feature change made no difference to performance relative to no feature changes. We therefore based our subsequent analysis on the performance to the cross-category feature change condition only and investigated whether there was an effect of the feature encoding modality on generalisation. To that end, we conducted a 3-way,

mixed subjects ANOVA with age group (4-6 years; 7-9 years; 10-13 years), learning modality (Visual; Haptic; Visuohaptic) and cross-category feature modality (visual feature; haptic feature; visuohaptic feature) on accuracy performance. For analysis of performance across the specific cross-category feature change (3x3x4) mixed design see Supplemental Materials 9.

As expected from the main findings, there was a main effect of age group [$F(2, 151) = 10.04$, $p < .001$, $\eta^2_G = .041$]. Bonferroni-adjusted pairwise comparisons revealed that children aged 7–9 years (M = 0.71, SD = 0.13, 95% CI [0.67, 0.75]) and 10–13 years (M = 0.72, SE = 0.11, 95% CI [0.68, 0.76]) performed significantly better than those aged 4–6 years (M = 0.59, SD = 0.18, 95% CI [0.53, 0.65]), with mean differences of [$t(151) = -3.85$, $p < 0.001$], and [$t(151) = -4.242$, $p < 0.001$] respectively. There was no significant difference between the performance of 7–9 and 10–13-year groups [$t(151) = -0.38$, $p = .92$].

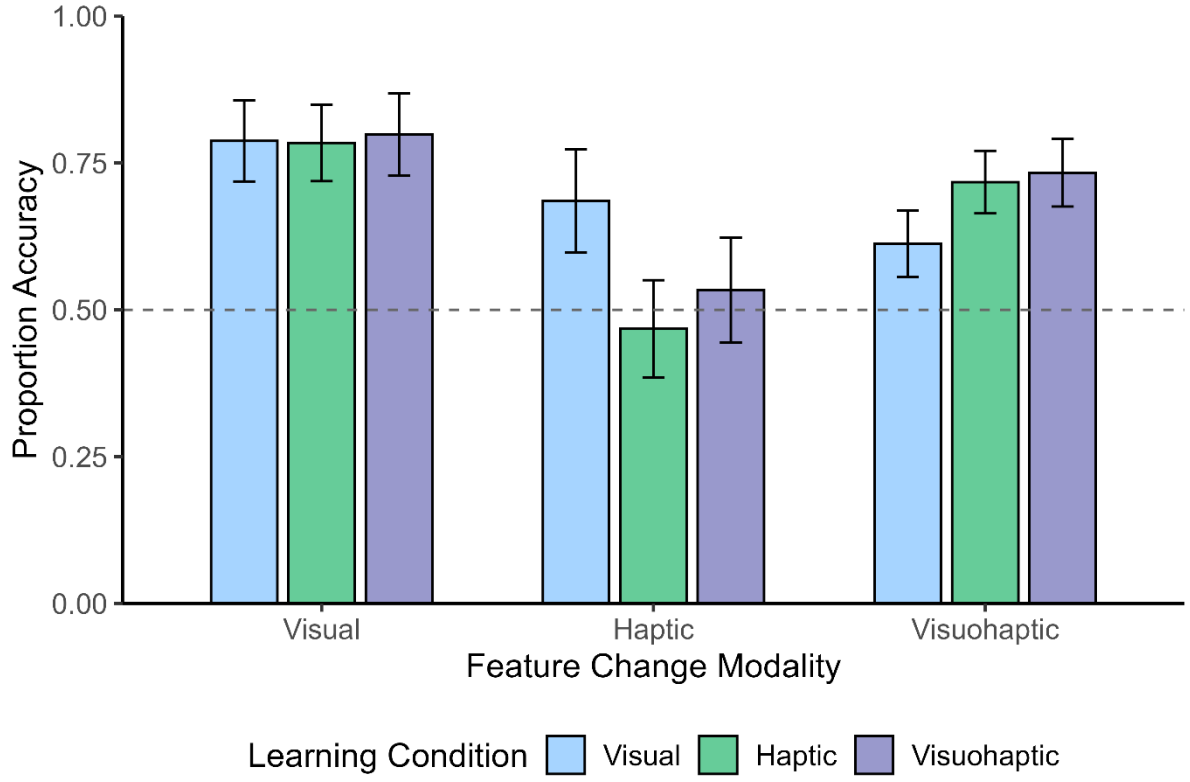
The main effect of learning modality was not significant [$F(2, 151) = 1.51$, $p = .224$, $\eta^2_G = .006$]. There was a main effect of the modality of the cross-category feature change [$F(1.77, 267.42) = 26.81$, $p < .001$, $\eta^2_G = .107$], indicating differences in generalisation accuracy depending on the sensory dimensions of the changed features. Bonferroni-adjusted pairwise comparisons showed that categorisation accuracy was significantly higher for objects in which the visual feature was changed (M = 0.78, SD = 0.25, 95% CI [0.73, 0.83]) than for objects in which either a haptic feature (M = 0.56, SD = 0.31, 95% CI [0.50, 0.62]; $t(151) = -6.77$, $p < 0.001$) or a visuohaptic (M = 0.68, SD = 0.21, 95% CI [0.64, 0.72]; $t(151) = -3.99$, $p < 0.001$) feature was changed across categories. Generalisation accuracy for a visuohaptic feature change was significantly greater than a haptic feature change ($t(151) = 3.76$, $p < 0.001$).

We found a significant, pairwise interaction between learning modality and the modality of the feature change [$F(3.54, 267.42) = 6.05$, $p < .001$, $\eta^2_G = .051$], which is shown in Figure 14. Follow-up Bonferroni-adjusted pairwise comparisons were conducted to compare the effect of the learning modality to each modality of the feature changes. First, taking generalisation performance to objects in which a visual feature was changed across categories, we found no significant differences across learning modalities (Haptic learners: M = 0.78, SD = 0.22, 95% CI [0.72, 0.85]; Visual learners: M = 0.78, SD = 0.26, 95% CI [0.70, 0.85]; Visuohaptic learners: M = 0.77, SD = 0.28, 95% CI [0.70, 0.84]; all $p > .99$). Next, taking generalisation to objects in which a haptic feature was changed, we found significantly better generalisation accuracy to visually (M = 0.70, SD = 0.26, 95% CI [0.62, 0.79]) than haptically (M = 0.47, SD = 0.31, 95% CI [0.39, 0.55]; t(151) = -3.91, $p < 0.001$) learned objects. Generalisation was also higher following visual than visuohaptic (M = 0.51, SD = 0.3395%, CI [0.42, 0.60]) learning (t(151) = 3.10, $p < 0.001$), and the difference between haptic and visuohaptic learning conditions was not significant ($p = 0.75$). For feature changes that occurred in both modalities (visuohaptic features) there were no differences found across learning modalities that survived correction for multiple comparisons ($p > 0.07$).

These findings align with the participants' qualitative responses with children who learned the categories using haptic and visuohaptic information reporting a greater dependence on the presence of haptic diagnostic features. In contrast, children who learned the categories using visual information only reported using visuohaptic features for categorisation (see Supplemental Materials 10 for more details).

Figure 14

Categorisation accuracy at test for objects with one cross category feature compared across learning modality condition



Note. Error bars represent 95% CIs and the dashed line represents chance performance.

Finally, we found no evidence for a two-way interaction between age group and the modality of the feature change [$F(3.54, 267.42) = 1.40$, $p = .233$, $\eta^2_G = .012$], indicating that the pattern of accuracy across feature modalities did not differ significantly by age group. The age group by learning modality interaction was also not significant [$F(4, 151) = 1.40$, $p = .238$, $\eta^2_G = .012$], nor did we find evidence for a three-way interaction between age group, learning modality and feature change modality [$F(7.08, 267.42) = 0.48$, $p = .848$, $\eta^2_G = .009$].

Discussion

The present study investigated the ability of children aged from 4 to 13 years to form novel object categories based on multisensory or unisensory information. Specifically, we asked whether children's ability to learn object categories, and flexibly generalise their

category learning to novel objects, differs depending on the modality used to acquire the category knowledge. Building on previous research (e.g. Broadbent et al., 2020; Gori et al., 2008), our initial aim was to assess whether children, in particular those in the group aged 10 years and older, show an advantage for learning if exposed to multisensory information relative to vision or haptic information alone. Moreover, we investigated the developmental trajectory the influence of the visual and haptic sensory systems on the formation of object categories. A second aim was to examine whether children can generalise their learning to objects with different features to their learned versions. We manipulated the modality (visual, haptic, or visuohaptic) of the cross-category features which were changed to help us identify the sensory basis of children's category representations.

Contrary to our initial expectations, we found no effect of modality on the efficacy of category learning (i.e. number of trials needed to reach learning criterion), nor did the learning performance vary across age groups. However, it is important to consider that success in learning the category task was nevertheless impacted by age: more children in the younger age group failed to learn relative to the other older age groups. Successful learners in the youngest group seemed to require more trials to learn in the visuo-haptic condition than in the other modalities, particularly vision. However, we did not find a benefit of visuohaptic over unimodal learning in any age group, as we had initially hypothesised. In doing so we do not replicate Broadbent's findings (2020) regarding children's category learning: they reported a benefit for unimodal haptic exposure in 6-year-olds and a benefit of either visuohaptic or haptic exposure over vision-alone in children aged from 8 years of age.

We then measured categorisation performance at test, to assess the nature of the category representations following learning. Our findings suggest that contrary to learning, categorisation performance at test was affected by age, with better performance in the older

children (i.e. from 7 years of age) irrespective of the learning modality. This finding suggests that generalisation performance to novel exemplars improves with development. This finding may be explained, at least in part, by the changes in sensory interactions with development. For example, it is suggested that crossmodal calibration, which benefits perceptual accuracy, is not achieved until about 8 years of age (Gori, 2015). Crossmodal calibration is a critical skill which allows the sense with the most direct access to the information to act as a calibrator for the sense which has indirect access. For example, holding an object using touch helps provide direct information about object size which is used to calibrate size in vision (Gori et al., 2010). Alternatively, the age-related effects may reflect improved cognitive ability including attention or memory and the retention of category knowledge at test. However, we also found a developmental effect due to the nature of the feature changes made, with worst performance when a feature change occurred across categories than within categories (or no change). The effect was particularly evident in the youngest child group. All participants performed this test visuohaptically; however, it may be that young children found it comparatively difficult to encode multiple features or perceptual redundancies. In support of this, only the 4–6-year-olds required more trials to reach learning criterion during the visuohaptic than either unimodal learning condition, and the visuohaptic condition may increase the number of features encoded (i.e. visual+haptic) or may make it less likely that multisensory features are selected from the other unimodal information.

To better understand the nature of the features used for categorisation, we focussed our analysis on the impact of cross-category feature changes in the generalisation task. We found that changes to visual features had little effect on categorisation performance across age groups and learning modality. Instead, haptic and visuohaptic cross-category feature changes were significantly more likely to affect categorisation accuracy across all learning

conditions. These findings are consistent with those of Broadbent et al., (2020) which suggested a dominance of haptic information when making category judgements. Although we found no evidence that haptic learning was easier, it seems that when haptic information was available during learning, either alone or in combination with vision, children selected and attended to this information for the purpose of categorisation and generalisation. However, it is important to note that the present design does not allow us to determine whether this effect is driven by haptic information itself, or by the way in which children selected or attended to features within the available sensory input. In particular, as visuohaptic features include visual information, it remains possible that children relied on the visual component of these features when forming category representations. This interpretation is supported by a large body of research demonstrating that vision often dominates multisensory perception when it provides reliable or salient information (Ernst & Banks, 2002), and that children's perceptual decisions may reflect modality appropriateness or task-specific weighting rather than stable modality dominance (Gori et al., 2008).

Interestingly, there was no observed interaction between age and the modality of the cross-category feature change suggesting that the utilisation of haptic features for categorisation was present across development. This finding is somewhat contrary to what we expected from previous work (Barutchu et al., 2009; Nardini et al., 2015; Nardini, Jones et al., 2008), that is, that older children may be more dependent on visuohaptic cross-category features and younger children on unimodal features, particularly vision (Misceo et al., 1999). However, given that children do not reliably integrate multisensory cues until late childhood and may instead rely on one modality at a time (Nardini et al., 2010; Gori et al., 2008), the absence of developmental differences may reflect similar modality selection strategies across age groups rather than equivalent multisensory representations.

However, differences in feature salience may have had an influence on categorisation, particularly if the haptic features were more distinctive than the visual features. Although differences in feature saliency cannot be ruled out in our study, it is important to note that each feature was equiprobable during learning and test, and that children were carefully instructed not to rely on one feature alone categorisation. Moreover, Broadbent et al. (2020) similarly found haptics to be the dominant modality for category learning in 6–7-year-olds. Nevertheless, future studies on categorisation might want to ensure that sensory information is equally discriminable across modalities.

It may also be important to consider the potential influence of the non-diagnostic (i.e. distractor) features (see Table 2) on categorisation performance. All features will have competed for children's attention, and the selection of the features may have been influenced by their relative saliency within each modality. For example, body colour (V) and compressibility (H) were non-diagnostic to category membership, but either of these features may have been more obvious than their corresponding within-modality features of leg colour or weight (which were diagnostic). Although we deliberately ensured that there was a balance between the diagnostic and non-diagnostic features in our task, it would be interesting to investigate whether a benefit for multisensory features on category learning would emerge if the relative number of non-diagnostic features was reduced in each object, thus decreasing the search space.

Though categorisation was highly successful across age groups (≥72% accuracy) it is clear from the learning phase that many participants, especially the youngest children, were unable to acquire these categories based on our object sets. Importantly, these objects were designed to be representative of the featural complexity of real-world objects: the objects children encounter everyday are complex, comprising of multiple features across modalities. However, future studies could help ensure that the task is more compatible with

children's cognitive abilities at each age. Efforts to reduce cognitive demands could include e.g. not presenting all diagnostic features at the same time in each trial but instead distributing these across trials or changing the ratio of diagnostic to distracting features. Moreover, the stimuli could be selected such that, across all age groups, the objects are equally discriminable prior to the experiment.

In summary, we found that children of all ages could successfully learn to categorise novel objects based on different encoding modalities. When tested on their ability to generalise the learned categories, performance suggested a featural dependency that was specific to haptic information. Given that the visuohaptic condition necessarily included visual input, it remains unclear whether observed effects attributed to visuohaptic, or haptic information were driven by haptic processing per se or by reliance on visual components within the multisensory input. Nevertheless, children could readily generalise their categories to new object exemplars that differed only in a visual feature, irrespective of which modality their category was acquired. However, they found it more difficult to categorise new object exemplars that differed in haptic featural information (unless they acquired the category using vision only). The present findings cannot disentangle the relative contributions of visual and haptic information within multisensory contexts, nor determine whether children integrated or selectively relied on one modality. Nonetheless, this research reveals the perceptual processes involved in categorisation in children and suggests that the formation of categories may be modality-dependent in all age groups.

Chapter 5

Visual Attention in Children and Adults during Visuohaptic Object Category Learning

Chapter 5

Abstract

Children's ability to categorise visual objects is evident from early stages of development although how categorisation is achieved based on multisensory information is unclear. Object categorisation relies on both top-down mechanisms, such as selective attention to relevant features or task goals, and bottom-up sensory encoding, which are all affected by developmental processes. This study examined whether task demands affect how children and adults attend to object features during category learning of novel objects, using eye-tracking and behavioural measures. Specifically, we assessed whether children (aged 8-11 years) and adults could adapt their feature search depending on prior knowledge of the test modality (visually-only, haptically-only) compared to no knowledge of the test modality. During learning, we measured fixation patterns to visual-only (e.g. colour) or visuohaptic features (e.g. texture) and whether these patterns were consistent with their knowledge of the test modality. We also tested categorisation and generalisation performance depending on their task instructions. Results showed that both age groups successfully learned the categories above chance. Adults outperformed children when prior visual information was available, whereas no age difference was found with prior haptic knowledge or unknown conditions. Participants who were not told their test modality learned in fewer trials, suggesting that uncertainty promoted more exploratory learning. Eye-tracking revealed that children produced a higher proportion of fixations overall, while both groups directed most fixations to visually salient features (even after controlling for area of interest size), indicating that bottom-up salience (e.g., colour contrast) guided attention more than task relevance. In the generalisation phase, participants with both visual and haptic prior exposure were more accurate than those with visual exposure alone, supporting evidence that multisensory experience enhances category representations. Together, the findings indicate that while visual salience strongly constrains attention,

multisensory experience and uncertainty can facilitate more efficient category learning and generalisation across development.

Introduction

Category learning frequently involves the integration of information across multiple sensory modalities. In everyday contexts, humans interact with objects not only through vision but also through touch, and the ability to flexibly integrate visual and haptic information develops over time and supports the formation of robust category representations. Theories of category learning have long emphasized the role of selective attention in shaping representations (Shepard et al., 1961; Gao et al., 2024). Forming novel object categories requires attention to and encoding of object features that are relevant to category membership, as well as the suppression or filtering of non-relevant information (Shepard et al., 1961). In theories of category learning such as the prototype model (Nosofsky et al., 1989; Smith & Minda, 1998), exemplar model (Medin & Schaffer, 1978; Nosofsky, 1986) and clustering model (Ashby et al., 1998) selective attention can be referred to as a 'weight' which is applied to stimulus dimensions during a categorisation judgement. Indeed, directing attention to category-relevant features, also known as category diagnostic features (Unger & Sloutsky, 2023), is critical. For example, if you were to attempt to categorise a mango as ripe you may check its colour or its firmness, shape in this instance is a non-diagnostic feature.

Developmental studies of visual category learning indicate that children distribute attention more broadly across stimulus features and are less likely to selectively weight diagnostic feature dimensions (Deng & Sloutsky, 2015; 2016). For example, 4-year-olds tend to rely on multiple probabilistic features even when deterministic ones are available, whereas older children and adults can shift attention to deterministic features during categorisation tasks (Deng & Sloutsky, 2015). Attentional cueing influences adults' and

older children's categorization and memory for features but has weaker effects on younger children, who show better memory for non-cued features, indicating more diffuse attention (Deng & Sloutsky, 2016). Blanco and Sloutsky (2019) found that adults exhibited greater costs of learned inattention when category rules shifted unexpectedly, whereas children's distributed attention allowed greater flexibility. These findings suggest that the development of category learning involves increasing selective attention and reduced reliance on distributed sampling.

The development of category learning involves a transition from early reliance on global similarity toward greater use of selective attention to object features with age (Sloutsky, 2003, 2010; Deng & Sloutsky, 2016). Younger children often adopt simple rules or rely on overall similarity, exhibiting diffuse patterns of attention to multiple stimulus dimensions (Huang-Pollock et al., 2011; Rabi et al., 2015). In contrast, adults can more efficiently select and attend to the most informative dimensions (Rehder & Hoffman, 2005), leading to faster and more accurate category learning (Deng & Sloutsky, 2016; Gao et al., 2024). Collectively, these findings suggest that children's attentional strategies during category learning are less strategic and more distributed than those of adults.

Most research on category learning has focused on unimodal (typically visual) tasks. However, everyday object recognition and categorisation frequently involves multisensory information, particularly vision and touch -consider reaching into your bag to locate your notebook. The development of category learning occurs across multiple sensory modalities, and the types of features available through each sense can shape how categories are represented. Object features can be either complementary or correlational across modalities, and both types can influence category formation (Newell et al., 2024). Certain features are unique to a single modality, for instance, colour is accessible only through vision and weight is perceived only through touch, whereas others, such as surface texture

and shape, are accessible through both vision and haptics, creating redundant multisensory cues (Jao et al., 2014). Even in early infancy, there is evidence for cross-modal equivalence: neonates demonstrate the ability to match objects across vision and touch based on features like shape and texture (Streri et al., 2000; Streri & Gentaz, 2004; Molina & Jouen, 1998). By around 2.5 years of age, children can reliably recognize objects using touch alone (Bigelow, 1981; Bushnell & Baxt, 1999), and this ability improves with age as haptic exploratory strategies become more efficient (Morrongiello et al., 1994). These findings indicate that multisensory information contributes to category learning from early in development, and that improvements in selective attention and sensory exploration may support a better use of both unimodal and multimodal object features for categorisation across childhood.

Cross-modal category learning provides an important context for understanding how attentional strategies and sensory processing interact during development. When object categories must be learned in one modality but retrieved or tested in another, learners must establish representations that generalize across sensory inputs. Neuroimaging studies have shown that visual and haptic object processing converge in higher-level visual areas (Amedi et al., 2001, 2005). For example, Jao et al. (2015) demonstrated that activation in the lateral occipital complex (LOC), a region associated with visual object recognition, was enhanced following vision-haptic cross-modal object exposure as compared to unimodal visual or unimodal haptic alone, with this cross-modal enhancement increasing from 7-8.5 years of age to adulthood. This finding suggests that the neural systems supporting object representations become increasingly modality-independent over development, potentially reflecting both maturation and experience-driven plasticity.

Behavioural studies reveal developmental differences in how children and adults deploy attention in cross-modal contexts. Coté (2015) examined visual attention during a

visual–haptic matching task and found that children and adults exhibited distinct fixation patterns, suggesting that children rely on different strategies for linking visual and haptic features. While adults often prioritize features that are diagnostic in the target modality (e.g., texture or shape for haptic recognition), children tend to rely more on visually salient or globally distributed features, which may be less effective for cross-modal transfer. These findings align with broader evidence that children exhibit more distributed attention during learning (Deng & Sloutsky, 2015; 2016), and that the ability to strategically select modality-relevant features improves with age. The efficiency with which children explore objects haptically continues to improve throughout childhood (Morrongiello et al., 1994), as does their ability to integrate complementary and redundant sensory information (Goi et al., 2012). Adults can flexibly adapt their visual attention to anticipate haptic demands, such as focusing on surface texture when they expect to later identify objects through touch. Children, by contrast, may not spontaneously adopt such strategies, resulting in less effective cross-modal transfer. Together, these findings suggest that developmental differences in cross-modal category learning may arise from both maturational changes in the neuronal architecture which subsumes multisensory integration as well as differences in strategic attentional control.

The present study examines how children and adults allocate visual attention to diagnostic object features during visual category learning under different anticipated test modality conditions, and how this attentional allocation relates to category learning efficiency, as well as subsequent cross-modal transfer and generalization performance. All participants first completed a visual category learning phase in which they were trained to categorise 12 now object stimuli, each defined along four diagnostic features: two visual-only features (body and leg colour) and two visuohaptic features (ear size and patch texture). Child and adult participants were randomly assigned to one of three test instruction

conditions: *Visual Test* – informed that they would later be tested visually*; Haptic Test* – informed that they would later be tested using haptics-alone; *Unknown Test* – informed that the test would be either visual or haptic (50:50). Following the learning phase, participants completed a categorization test in either the visual or haptic modality, using the same 12 prototypes. Finally, all participants conducted a second learning session and then completed a generalisation test, which was visual-only, involving 24 novel objects constructed with a family resemblance structure, which required extending learned category knowledge to new exemplars.

Eye movements were recorded during the learning phases and generalisation test to examine attentional allocation to different feature types. On each object, Areas Of Interest (AOIs) were identified which corresponded to the location of the four diagnostic features, allowing calculation of both the proportion of fixations and fixation durations directed to visual-only versus visuohaptic features. Behavioural performance was measured in terms of proportion correct during both the categorization and generalization tests. We hypothesized that adults would be more efficient at selectively attending to the task modality-relevant features during learning than children. In turn, this strategic allocation of attention during learning was expected to support better accuracy in the categorization test. In contrast, we hypothesized that children would exhibit more distributed attention across all features, reflecting less strategic and more exploratory learning. Consequently, children were expected to show lower categorisation accuracy in the categorization task compared to adults.

Method

Participants

We recruited 86 children (38 male) to take part in this study from five primary schools (children in 2nd-4th class) located in the greater Dublin area. All 65 adult participants

(26 male) were recruited from Trinity College Dublin. An a priori power analysis conducted using Pangea (Westfall, 2015) determined that a sample size of 120 participants overall (20 per age x instruction condition) would be required to detect interactions with a medium effect size (*f*) of 0.25, $a = 0.05$, and power $(1- \beta) = 0.80$. For the purposes of our experiment, the participants were grouped according to age with one child (mean age = 9.14 years, age range = 8-11 years) and one adult (mean age = 25.10 years, age range = 19-32 years) group. These groups were then further allocated into one of three instruction conditions: a visual test, a haptic test, either a visual or haptic test (referred to as the 'unknown' condition); see table 7 for details of the number of participants per condition. Recruitment continued until we attained a minimum of 20 participants per age group per instruction condition who successfully completed all phases of our experiment.

Ethical approval for the experiment was given by the School of Psychology Research and Ethics Committee at Trinity College Dublin (Approval Number: 3782). Consent was acquired in written format from the adult participant and from the parent/guardian of the participating children. Assent was also acquired verbally from each child prior to study commencement. Adult participants were furnished with an information sheet at least a week prior to testing. The information sheets and parental consent forms were distributed by classroom teachers and principals to parents or guardians. All participants reported normal or corrected to normal vision.

Table 7

Number of participants in each age group allocated to each instruction condition

Instruction Condition	Age Group		
	Child	Adult	**Total**
Visual	28	24	52
Haptic	27	21	48
Unknown	31	20	51
Total	86	65	151

Stimuli and apparatus

The stimuli used in this experiment were adapted from Chapter 4 of this thesis and the basic structure of object stimuli described by Broadbent et al. (2020) and were made to resemble two novel alien species. These alien species were named ‘Dorps’ and ‘Neems’ and were utilised in this experiment as our two novel object categories. Objects were made of four different fabric ‘beanbag’ shapes -an ovoid, a hemisphere, a triangular based pyramid and a square based pyramid. Additional object features were attached to these fabric shapes: hair; four legs; patches; ears; eyes. The size of the objects was standardised so that their maximum dimensions were 10cm height x 10cm width x 5cm depth. As such they could be easily handled by child participants across the different ages. We manipulated the stimulus features to create two object categories. To do this, we defined four object features as diagnostic of category membership. These features included two visual object features and two which were visuohaptic i.e. they are accessible to both the visual and haptic sensory modalities. Further details are as follows:

Visual-only feature 1: The object leg colour was either warm (orange/yellow) or cool (purple/blue). Each object had four legs (two on each side) and these were the same colour as each other.

Visual-only feature 2: The object body colour was either warm (orange/red) or cool (light blue/dark blue). All objects were made of the same type of material (cotton).

Visuohaptic feature 1: The object patches which were affixed to the body, each object had seven 'patches' or rectangles, each measuring an area of 1cmx3cm, which were equally distributed across the surface of the body. Different textures were used for the material of the patches which were either smooth (ribbon or felt) or rough (plastic mesh or hessian).

Visuohaptic feature 2: The object ‘ears’ consisted of four variants of buttons, these buttons were matt white and ranged from 10 to 20 mms. The ‘ears’ fell into two categories

small (10mm and 12.5mm) to large (17.5mm and 20mm) and were positioned below the eyes (to appear as ears) and were located along the object seams on the left and right between the two legs on each side.

The objects had additional features which we termed ‘distractor’ features. In order to balance the number of cues available in each test condition we introduced cues which were accessible only by the haptic modality which were object weight and compressibility. There were an additional four features which functioned as visuohaptic distractors. These included the object shape, leg material texture, hair type and eyes. These distractors were non-diagnostic of category membership and were randomly distributed across the object categories. For more details of object composition and the distribution of object features across each stimulus see Supplemental 11.

Of the object stimuli (24 objects), half of these objects belonged to Category A and the other to Category B. From the entire set, 8 objects were used for the category learning phases, and these objects had all four diagnostic features for category membership (see Figure 15 for details). The categorisation phase included these 8 objects as well as a further 16 objects (24 in total) which were used to test generalisation. Of these 16, half belonged to Category A and half to Category B. Category membership for these 16 novel objects was based on a ‘family resemblance’ design. This means that of the 4 diagnostic features used for category membership, 3 were original and one was replaced by a feature from the other category. The type of cross-category feature used was counterbalanced across stimuli, with both variants presented i.e. if bright legs are a diagnostic feature for the ‘Dorp’ category, at generalisation test there would be two Dorps with the leg cross-category feature: one with purple legs; one with blue legs, and vice versa for the Neem object category.

During the learning stage, images of the objects were used, ensuring that all diagnostic and distractor features were visible. To create the images, each object was

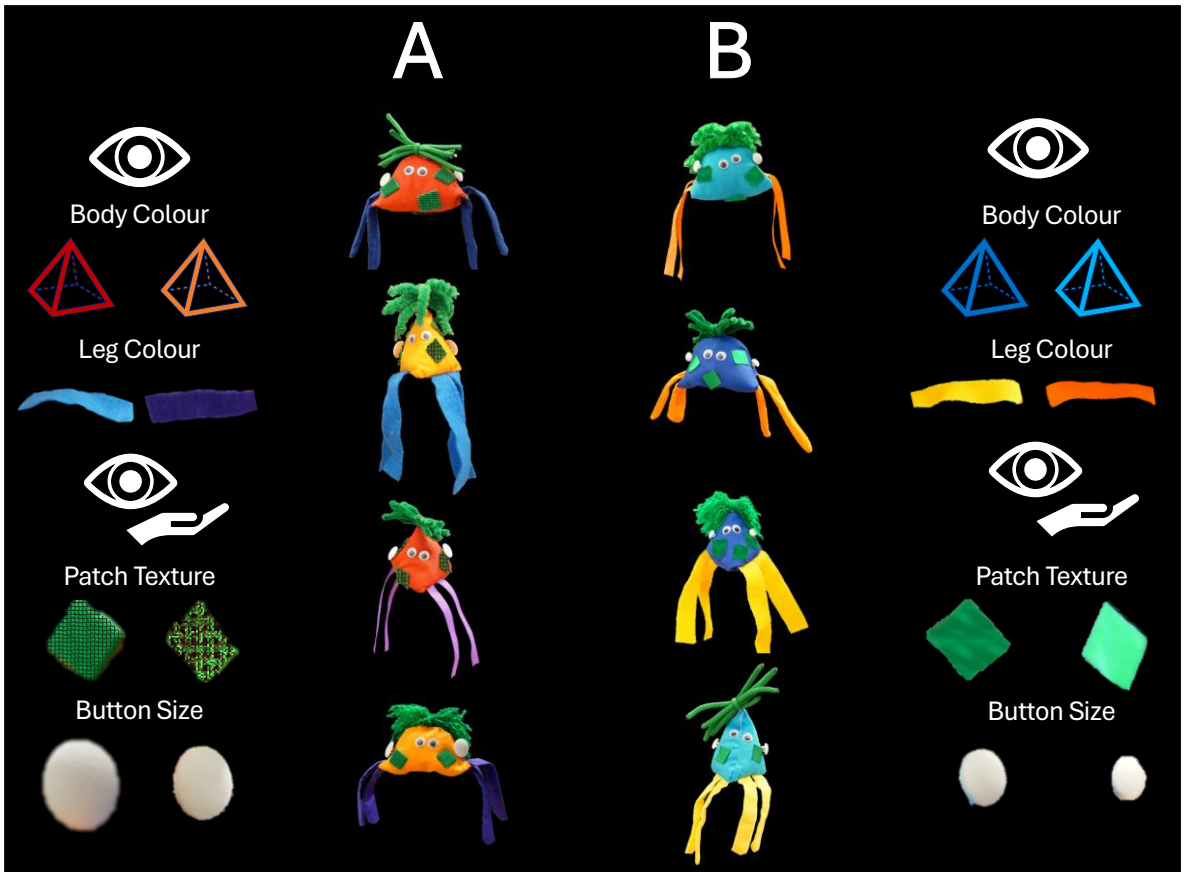
photographed using a Nikon camera at a distance of 30cm. The objects were positioned on a stand and at a 25° angle to standardise the visual stimuli and ensure all object features were in frame (see Supplemental 12 for details). A black fabric was used as a background.

All testing took place in either the hosting school or the dedicated laboratory space at Trinity College Dublin. A Tobii Fusion eye tracker, positioned at the bottom of a computer screen, was used to measure fixation number and duration throughout all category learning phases (as well as in our final generalisation test). The experiment stimuli were presented and the data recorded, both behavioural responses and eye tracking data, using Psychopy 2025.2.0. We used a laptop (Alienware 15 R4) with a 40cm (resolution of 1920 by 1080 pixels), 60 Hz IPS display screen to display the images during the learning sessions. This laptop was positioned approximately 60cm away from the participant when they were seated at a desk. The participants used a keyboard with stickers on the appropriate keys to make a category response ‘d’ for Dorp and ‘n’ for Neem. Each participant wore headphones to receive auditory feedback during learning trials. Eye tracking calibration was first conducted using Tobii Pro Lab Manager software, with a validation routine added within the coded experiment on Psychopy to ensure correct recording. This calibration was repeated as required throughout the experiment prior to the start of each phase.

For participants who completed the categorisation test in the haptic modality, they were presented with each object via an occlusion box. This box allowed the experimenter to place and remove each object out of view of the participant. As the participant’s hands were used to explore the objects, the experimenter inputted their responses via a keyboard after the participant verbally responded their category choice.

Figure 15

Learned object stimuli and their category diagnostic features relevant to each of the two categories



Note. See isolated category diagnostic features for category A (e.g. 'Dorp') on the left and category B (e.g. 'Neem') on the right of the panel. Visual features are indicated by an eye symbol and visuohaptic features are indicated by a hand-eye symbol. All 8 learned objects were presented in learning 1, category test 1 and in learning 2 phases of the experiment.

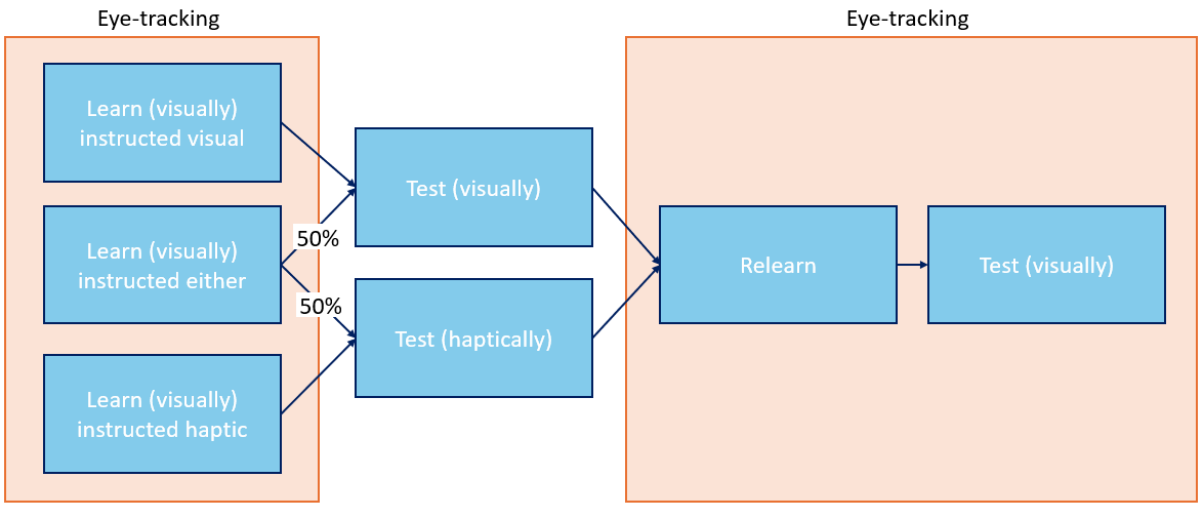
Design

This experiment was based on a fully factorial between subjects' design with two between subjects' factors. These factors were age group: child versus adult; and instruction condition i.e. the modality to be used at test: visual, haptic, unknown modality. To explore how participants allocate attention to the object stimuli, we utilised eye tracking during the (visual-only) learning blocks as well as in the final generalisation test. Eye tracking was not possible during the first test phase due to haptic and visual object exploration. There

were four phases to the experiment: a first learning phase, test phase, second learning phase and generalisation test (see figure 16 for illustration).

Figure 16

Experimental procedure with both learning and test phases with periods of eye-tracking use outlined



The first learning phased included a basic set of 8 trials each demonstrating a unique object, and each trial was repeated a minimum of 3 times and a maximum of 8 times to ensure learning (i.e. a range of 24-64 trials). Participants were required to achieve a learning criterion 75% correct and following successful learning they then completed the first categorisation test. This test took place in either the visual or haptic modality, depending on instructions. Participants allocated to the 'unknown' instruction condition conducted the test either in the visual or haptic modality (this was counterbalanced across age groups to ensure equal sample sizes for each test modality). The categorisation test included 16 trials, comprising the basic set of 8 trials (presenting each of the learned objects) which was repeated once. The participant then underwent a second learning phase, which was added to ensure that the children, in particular, remembered the object categories and refresh their learning by receiving feedback on their responses once more. Each of the 8 trials were repeated a minimum of twice with the maximum of 8 repetitions in this second phase. In

the final phase, participants underwent a generalization test, in which all 24 stimuli were presented. These stimuli included the learned objects and 16 novel objects which were designed based on family resemblances (see Stimuli). Eye movements were again recorded in the generalisation phase to identify the features that were attended to for task performance.

To derive eye tracking measures, we first defined 11 rectangular areas of interest (AOIs) that encompassed the entire object stimulus. These AOIs were: the left and right eye, left and right legs, body, patches 1-3, hair and the left and right ear; these AOIs were then coded as either a *vision-only* feature (body and left and right leg), a *visuohaptic* feature (patches 1-3 and left and right ear) or as *other body feature* (left and right eye as well as hair). All fixations which fell beyond the bounds of the AOIs were discarded, as was the first 500ms per trial (when the fixation cross appeared on screen) and for learning phases any time after 6 seconds (when the stimulus was no longer on the screen) and after 2 seconds in the generalisation phase (when the stimulus was no longer on the screen). Our dependent variable of interest *proportion of fixations* was calculated per AOI type per trial as follows:

$$proportion\ of\ fixations = \frac{\text{the number of fixations per AOI region}}{\text{total number of fixations within AOI regions}}$$

A second dependent variable of interest *mean duration of fixations* was also selected with the total duration of fixations per AOI calculated per trial and averaged across the final block of trials per participant.

Procedure

To take part in the experiment, each participant was seated at a table with the laptop screen located approximately 60cm away from them. Prior to the experiment, the participant completed the Tobi Eye tracker 5-point calibration while minimising their blinking. They were verbally told to maintain their position and were prompted again

during the task if they began to lean or move. They then read the task instructions outlining the purpose of the experiment: to learn to tell the difference between two types of alien creature. To do so they needed to press either the ‘d’ (‘Dorp’ category) or ‘n’ (‘Neem’ category) key on the keyboard (category labels were counterbalanced across participants). The keys were indicated with stickers. Participants were required to hold their index fingers over the relevant keys during the task.

Participants were required to learn to sort two types of aliens and were instructed that they would then complete a test either visually- just like how they are learning them, using haptics alone- where they were shown the occlusion box and informed they would be feeling the objects, or that they may be tested visually or haptically. They then underwent an eye tracking validation routine to ensure they were seated correctly with the laptop aligned with their gaze. Participants were told that randomly, throughout the task, they would encounter asteroids as part of our mini game, when presented the asteroid they would need to look at it. In addition to maintaining attention, this mini-game was inserted within each block of trials to ensure the eye tracker was calibrated with their gaze (i.e. that participants were able to fixate on the asteroid when presented in a known location). If there was misalignment, the experiment would not proceed until the participant repeated the calibration routine.

The initial learning task included at least three repetitions (blocks) of all learned stimuli. Accuracy was measured per block and after the minimum number of 24 trials had been completed (3 blocks) the participant could proceed if their accuracy reached at least 75%. A learning trial proceeded as follows: fixation cross was centrally located on the screen for 250ms followed by an image of an object presented for 6 seconds. The participant was able to respond after 2s had elapsed whilst the object image was displayed. There was no time limit to the responses, however, the object image was removed from the

screen after 6s. Feedback in the learning phases was both a 500ms visual message of 'correct' or 'incorrect' as well as auditory feedback directly following the participant's response in each learning trial. Participants who did not successfully reach the learning criterion of accuracy -after 8 repetitions- were informed they had completed the experiment and thanked for their time.

For successful learners, the first test block followed the learning and was conducted in either the visual or haptic modality. Participants assigned to the 'unknown' condition were informed of the test modality just before the test, and testing proceeded in the same manner as the other conditions. The test consisted of 16 trials, and the participant could respond from the start of stimulus presentation. There was no feedback given to the test trials. For the haptic test, the participants used an occlusion box to explore the objects without being able to see the objects. They verbally reported the stimulus category to the experimenter who then recorded their responses on the laptop.

Once this first test phase was complete, all participants then completed a second calibration routine to ensure accurate eye tracking. The participants then underwent a second learning phase with the same objects as the first learning phase, to 'refresh' their learning. Following successfully reaching the learning criterion in this second learning phase, the participants underwent a final test phase to assess generalisation. This block consisted of 24 trials (one per unique object) with one repetition and was conducted in the visual modality only. The participants were able to respond as soon as the object image appeared in a trial, and no feedback was given. All participants were then debriefed at the end of the experiment. In total, the experiment took between 20-25 minutes per participant.

Results

Learning Phase 1: Participants' performance

In the course of the first category learning phase 13 participants (8 children and 5 adults) failed to reach the learning criterion and did not continue with the experiment. This led to a final sample size of 138 participants. Of the child participants (M = 9.62, SD = 0.52; 4 males) who did not reach the 75% accuracy criterion, 1 was allocated to the visual test, 4 to the haptic test and 3 to the unknown testing modality. In the adult group, 4 were allocated to the visual test, and 1 to the haptic test conditions (M = 27.4, SD = 2.88; 2 males). Further details on the final participant numbers per age group and instruction conditions are provided in Table 8.

Table 8

Details on age of the participant groups who reached the learning criterion in the first learning phase

Age Group	Instruction Condition	N	*M (years)*	*SD (years)*	Min age	Max age	% Males
Child	Haptic	23	9.15	0.88	8	10	40
Child	Unknown	28	9.18	0.96	8	11	45
Child	Vision	27	9.12	1.03	8	11	46
Adult	Haptic	20	25.0	3.50	19	30	50
Adult	Unknown	20	25.6	2.75	20	31	28
Adult	Vision	20	24.8	3.47	20	32	37

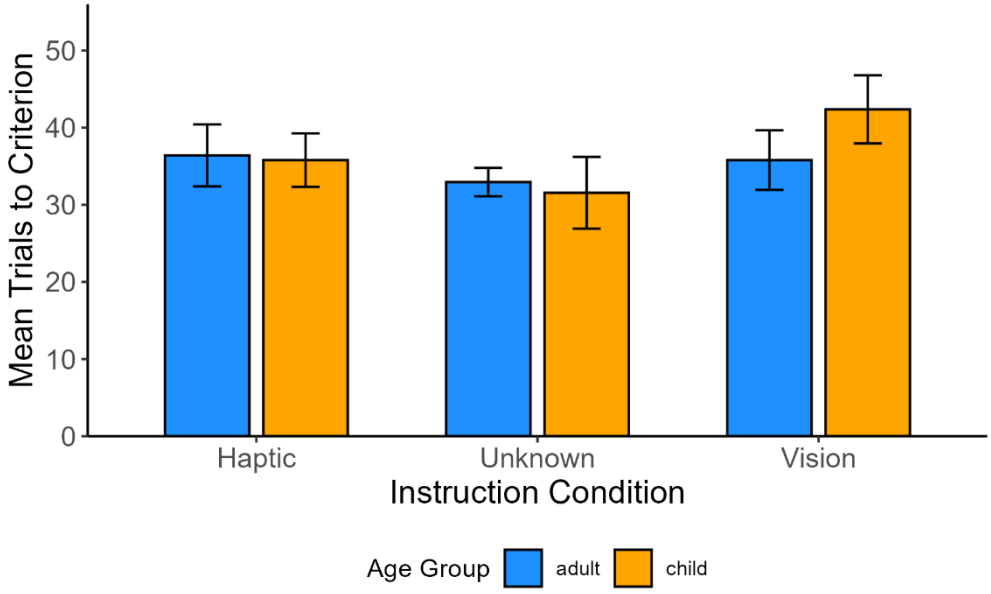
Learning Phase 1: mean number of trials to successful learning

To reach the category learning criterion, participants were required to complete a minimum of 3 blocks of 8 trials leading to a minimum trial number of 24. We wished to compare the efficacy of category acquisition across task instruction conditions (visual test, haptic test, and unknown test modality) and age groups (children and adults). However, the data violated assumptions of normality and homogeneity of variance. Specifically, the Shapiro-Wilk test indicated that the residuals were not normally distributed, W = 0.867, $p < .001$, and Levene's test revealed a significant violation of homogeneity of variance [$F(5, 114) = 3.29$, $p = .008$]. Due to these violations, a 2x3 aligned rank transform (ART) ANOVA

was conducted. The ART ANOVA revealed no effect of age group [$F(1, 114) = 3.03$, $p = .084$, $\eta_p^2 = 0.03$]. However, there was a main effect of instruction condition [$F(2, 114) = 7.40$, $p < .001$, $\eta_p^2 = 0.11$] with fewer trials required to reach the learning criterion in the 'unknown' testing condition ($M = 32.2$. $SD = 7.64$, 95% CI [29.7, 34.8]) compared to the visual test condition ($M = 39.7$. $SD = 10.9$, 95% CI [36.5, 42.8]; [$t(114) = -3.84$, $p < 0.001$]). No significant differences were observed between the number of trials required to reach criterion in either the 'unknown' or 'visual' conditions compared with the haptic test instruction condition ($M = 36.1$, $SD = 8.38$, 95% CI [33.5, 38.7]). A significant interaction was observed between age group and instruction condition [$F(2, 114) = 3.77$, $p = .026$, $\eta_p^2 = 0.06$] as shown in Figure 17. The post hoc pairwise comparisons with Bonferroni correction revealed a significant difference between the number of blocks taken to reach the learning criterion between children and adults in the visual instruction condition, with adults taking ($M = 35.8$, $SD = 8.59$, 95% CI [31.9, 39.7]) significantly fewer trials than children ($M = 42.4$, $SD = 11.7$, 95% CI [38.0, 46.8]; [$t(114) = -2.53$, $p = 0.01$]) in this condition only. There were no differences between the age groups for the haptic or 'unknown' test groups.

Figure 17

Number of trials required to successfully pass the category learning task for children and adults across the different test instruction conditions



Note. Error bars represent 95% confidence intervals (CIs).

Categorisation Test Phase 1: accuracy performance

As testing could be performed under visual or haptic conditions, analyses of the test phase included the additional factor of Test Modality. A linear mixed effects model using the lme4 package (Bates et al., 2015) was conducted to examine the effects of Age Group (child vs adult), Testing Modality (Haptic vs Vision), and instruction condition (Known vs Unknown) as well as their interactions on categorisation accuracy at test. Prior to analysis, the assumptions of normality and homogeneity of variance were assessed. The residuals were approximately normally distributed [$W(128) = 0.98$, $p = .052$], and the assumption of homogeneity of variance was satisfied [$F(7, 120) = 1.86$, $p = .082$], as indicated by Levene's test. The model explained approximately 22% of the variance in mean accuracy, $R^2 = .22$, adjusted $R^2 = .19$.

There was a significant main effect of *age group* [$F(1, 120) = 17.95$, $p < .001$, $\eta_p^2 = .13$]. Across both modalities the adults ($M = 0.81$, $SD = 0.22$, 95% CI [0.75, 0.86])

outperformed children (M = 0.65, SD = 0.23, 95% CI [0.59, 0.70]), with a mean difference of 0.13 [$t(120) = 3.30$, $p = .001$] (Bonferroni-adjusted). A significant main effect of *testing modality* was observed [$F(1, 120) = 5.43$, $p = .021$, $\eta_p^2 = .04$], with higher accuracy in the vision condition (M = 0.76, SD = 0.24, 95% CI [0.70, 0.81]) compared to haptic (M = 0.68, SD = 0.23, 95% CI [0.62, 0.74]). There was no significant effect of instruction condition [$F(1, 120) = 2.84$, $p = .09$, $\eta_p^2 = .02$].

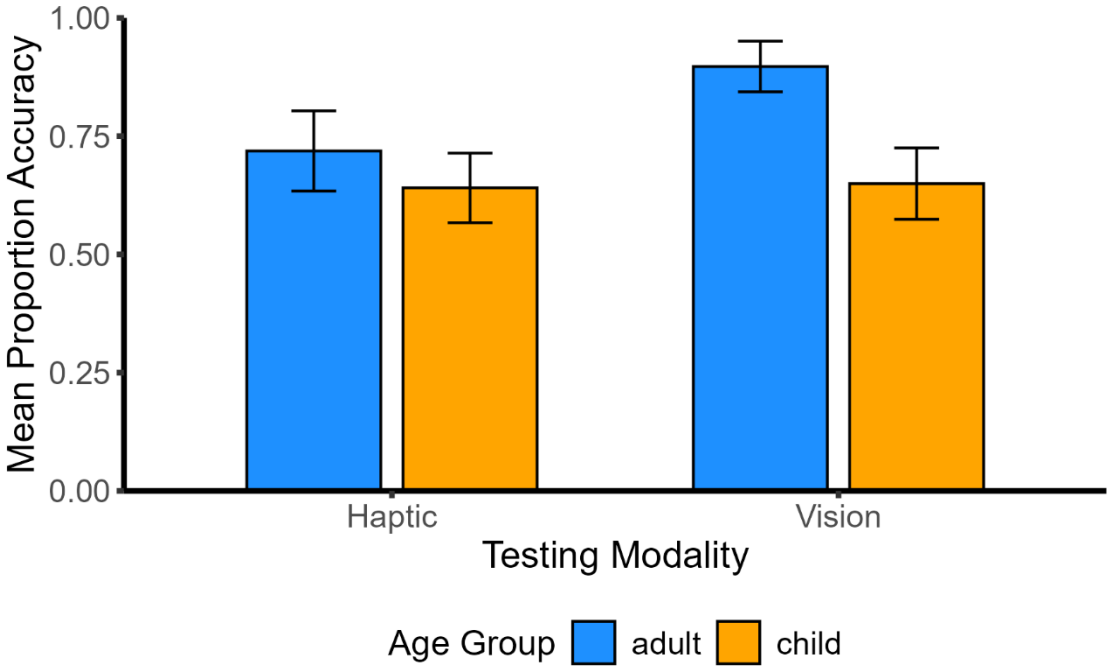
There was a significant interaction between age group by testing modality [$F(1, 124) = 4.63$, $p = .03$, $\eta_p^2 = .04$] as shown in Figure 18. Post hoc analyses revealed that in the haptic condition, adults ($M = 0.72$, $SD = 0.24$, 95% CI [0.63, 0.80]) and children ($M = 0.64$, $SD = 0.21$, 95% CI [0.57, 0.71]) did not differ significantly [$t(120) = 0.94$, $p = .347$]. In contrast, in the vision condition, adults ($M = 0.90$, $SD = 0.15$, 95% CI [0.84, 0.95]) performed significantly better than children ($M = 0.65$, $SD = 0.24$, 95% CI [0.57, 0.73]; [$t(120) = 3.74$, $p < .001$]). The two-way interaction between age group and instruction condition [$F(1, 120) = 3.78$, $p = .05$, $\eta_p^2 = .03$], and between instruction condition and testing modality [$F(1, 120) = 0.0003$, $p = .99$, $\eta_p^2 < .001$], did not reach significance. Nor was the three-way interaction between age-group, test modality and instruction condition [$F(1, 120) = 0.01$, $p = 0.92$, $\eta_p^2 < .001$]. These results indicate that the age advantage in categorisation accuracy was primarily driven by the vision testing modality, with adults performing better than children although children's performance was above that expected by chance.

Figure 18

Categorisation accuracy in the visual or haptic modality compared across children and adults

Learning Phase 2: Participants' performance

Four children failed to reach the learning criterion threshold leading in the second learning phase, reducing the overall sample size to 134 participants. Of these participants



(M = 10.2, SD = 0.5; 2 males), 3 had been tested using vision-only (n = 3) and 1 was tested haptically (n = 1). Further details on participant numbers of successful learners per age group and previous test modality are provided in Table 9.

Table 9

Details on the age of the participant groups who reached the learning criterion in Learning Phase 2

Age Group	Instruction Condition	N	*M* *(years)*	*SD* *(years)*	Min age	Max age	% Males
Child	Haptic	22	9.03	0.94	8	11	41
Child	Unknown Haptic	13	8.91	1.04	8	11	55
Child	Unknown Vision	13	9.45	0.82	8	11	60
Child	Vision-only	26	9.19	0.98	8	11	42
Adult	Haptic	20	25.2	3.17	19	30	43
Adult	Unknown Haptic	10	26.2	2.33	24	31	67
Adult	Unknown Vision	10	24.9	3.10	20	30	77
Adult	Vision-only	20	25.0	3.22	20	32	33

Learning Phase 2: mean number of trials to successful (re)learning

In order to progress to the final generalisation phase, we repeated the learning phase in order to refresh participant category knowledge. The object stimuli were the same as in Learning phase 1, however, all participants either had previous experience of the objects using vision only or also had haptic experience, depending on their task allocation in phase 1. Moreover, the participants previously allocated to the 'unknown' modality were divided according to whether they were tested visually or haptically. In this second learning task participants were required to complete a minimum of 2 blocks of 8 trials leading to a minimum number of trials of 16.

Here we wished to assess if participants with haptic experience led to more rapid learning in this second phase task than those with visual experience only and whether prior knowledge of the test modality affected performance across age groups (children and adults). However, the data violated assumptions of normality and homogeneity of variance. Specifically, the Shapiro-Wilk test indicated that the residuals were not normally distributed, [$W(120) = 0.74$, $p < .001$], and Levene's test revealed a significant violation of homogeneity of variance [$F(7, 112) = 3.10$, $p = .005$]. Due to these violations, as well as consideration to the count data nature of the dependent variable trials to learning criterion, a nonparametric factorial analysis was conducted using the aligned rank transform (ART)

procedure with age group (child, adult) and prior modality of exposure (visual (learning and test), visual-learning and haptic-test) as between-subjects factors and instruction condition (known, unknown) as a within-subject factor.

The 2x2x2 ART ANOVA revealed a main effect of age group [$F(1, 112) = 23.79$, $p < .001$, $\eta_p^2 = 0.18$], with adults ($M = 17.9$, $SD = 6.65$, 95% CI [16.1, 19.6]) taking significantly fewer trials to reach learning criterion than the children ($M = 21.7$, $SD = 9.78$, 95% CI [16.1, 19.6]; [$t(116) = --4.877$, $p < 0.001$]). The effect of modality of prior exposure failed to reach significance [$F(1, 112) = 3.86$, $p = .052$, $\eta^2_p = .03$]; but displayed a trend for fewer trials required for those with visual and haptic prior exposure ($M = 19.2$, $SD = 8.34$, 95% CI [17.00, 21.3]) compared to visual only exposure ($M = 23.4$, $SD = 13.8$, 95% CI [19.9, 26.8]). No significant effect of instruction condition was observed [$F(1, 112) = 0.11$, $p = .741$]. The interactions between age group and modality of prior exposure [$F(1, 112) = 1.64$, $p = .204$], age group and instruction condition [$F(1, 112) = 0.81$, $p = .370$], and between modality of prior exposure and instruction condition [$F(1, 112) = 3.11$, $p = .081$] did not reach significance; nor did the three-way interaction [$F(1, 112) = 3.39$, $p = .068$].

Test Phase 2: participants' categorisation performance to learned objects

An initial check of participant accuracy at second test revealed that 10 participants performed below chance when categorising previously learned objects, therefore their data were removed from further analysis. These participants included 9 children ($M = 8.67$, $SD = 0.71$; 7 boys) and 1 adult leaving data from 124 participants for further analysis. Details of the final number of participants in each of the learning conditions is shown in Table 10, with a full illustration of participant counts at each experimental phase in Supplemental 13.

Table 10

Details on age of the participant groups who reached the generalisation task compared across modality of prior exposure

Age Group	Prior experience	N	*M (years)*	*SD (years)*	Min	Max	% Males
Child	Haptic and Vision	34	9.08	0.96	8	11	67
Child	Vision-only	31	9.45	0.95	8	11	45
Adult	Haptic and Vision	29	25.1	3.21	20	30	41
Adult	Vision-only	30	25.0	3.17	20	32	33

An initial assessment of accuracy to repeated objects (learned stimuli) was conducted with age-group (child versus adult) and modality previously exposed (vision only, visual and haptic) as factors. A linear mixed-effects model was fitted using maximum likelihood estimation and the lme4 package (Bates et al., 2015), with prior exposure modality (vision (learning and test), visual-learn and haptic-test) and age group (child, adult) as fixed effects, their interaction, and a random intercept for participant to account for repeated measures. The model explained a moderate proportion of variance, with a marginal $R^2 = 0.14$ (variance explained by fixed effects) and a conditional $R^2 = 0.58$ (variance explained by the full model, including random effects) indicating that approximately 50% of the total variance was attributable to differences between participants (ICC = .51).

Type-III Wald χ^2 tests revealed a main effect of prior exposure modality [$\chi^2(1) = 6.16$, $p = .010$], with higher categorisation accuracy to visual and haptic exposure ($M = 0.91$, $SD = 0.11$, 95% CI [0.87, 0.94]) relative to visual exposure alone ($M = 0.86$, $SD = 0.14$, 95% CI [0.83, 0.90]). There was no effect of age group [$\chi^2(1) = 1.56$, $p = .21$]. Finally, there was no significant prior exposure by age group interaction, $\chi^2(1) = 2.34$, $p = .13$. Overall, these results suggest that categorisation performance of all ages benefitted from prior crossmodal exposure relative to visual only exposure to the objects.

Test Phase 2: participants' generalisation performance to novel objects

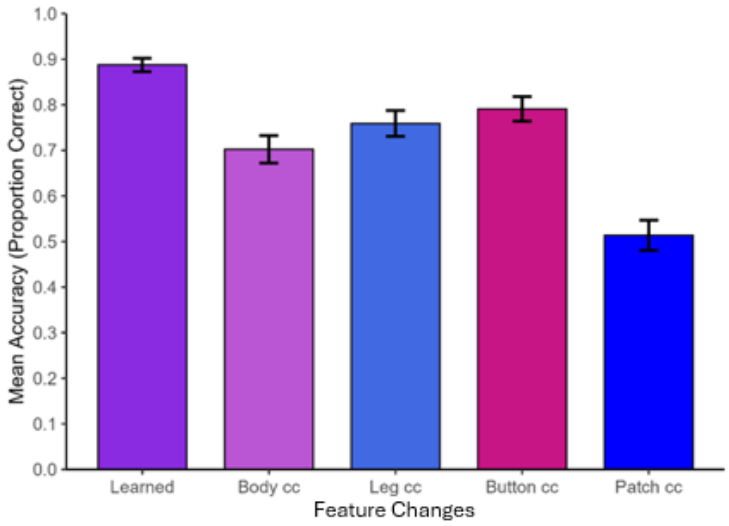
To examine whether accuracy for novel objects varied as a function of age or prior experience, we also examined whether the object feature type affected performance. Here feature type refers to the introduction of cross category features in our novel objects; if substituting a category diagnostic feature with that of the opposing category results in decreased accuracy, this indicates that that specific feature acts as the primary feature used in category judgement. Feature type was treated as a factor in the model with five levels: learned objects (which functioned as the reference condition in our model); cross category body colour ($body_{cc}$); cross category leg colour (leg_{cc}); cross category button size ($button_{cc}$); cross category patch texture ($patch_{cc}$). Our other fixed factors were modality of prior exposure (vision only vs. visual and haptic), and age group (child vs. adult). We fitted a linear mixed-effects model using maximum likelihood estimation and the lme4 package (Bates et al., 2015). The model included random intercepts and random slopes for feature type by participant to account for individual variability in accuracy patterns across object feature types. Model comparison using a likelihood ratio test showed that including random slopes significantly improved model fit relative to a random-intercept-only model [$\chi^2(14) = 1720.8$, $p < .001$]. Model performance indicated that the fixed effects explained a modest proportion of variance, with a marginal $R^2 = .12$ with random effects accounting for 48% of the variance observed in our model conditional $R^2 = 0.48$.

Type-III Wald χ^2 tests revealed a significant main effect of object feature type [$\chi^2(4) = 114.86$, $p < 0.001$]. Post-hoc tests indicated that accuracy was significantly higher for the learned features ($M = 0.89$, $SD = 0.32$, 95% CI [0.87, 0.90]) than for those with $body_{cc}$ ($M = 0.70$, $SD = 0.46$, 95% CI [0.67, 0.73]; $t(98.3) = 5.31$, $p < 0.001$); leg_{cc} ($M = 0.76$, $SD = 0.43$, 95% CI [0.73, 0.79]; $t(97.9) = 4.05$, $p = 0.001$); $button_{cc}$ ($M = 0.79$, $SD = 0.41$, 95% CI [0.76, 0.82]; $t(97.8) = 3.40$, $p < 0.01$) or $patch_{cc}$ ($M = 0.51$, $SD = 0.50$, 95% CI [0.48, 0.55]; $t(98.5) = 9.41$, $p < 0.001$). Additionally, a significant difference was found between

the $body_{cc}$ and the $patch_{cc}$ [$t(98.6) = 3.00$, $p = 0.03$] feature changes; and the leg_{cc} was also more accurate than $patch_{cc}$ feature change [$t(98.5) = 9.41$, $p < 0.001$]; the $button_{cc}$ was more accurate than $patch_{cc}$ feature change [$t(98.5) = 5.35$, $p = 0.001$]. See Figure 20 for details.

Figure 19

Main Effect of object feature changes



Note. Error bars represent 95% confidence intervals; cc refers to objects with cross category features, with $Body_{cc}$ and Leg_{cc} referring to visual diagnostic features and $button_{cc}$ and $Patch_{cc}$ referring to the visuohaptic diagnostic features.

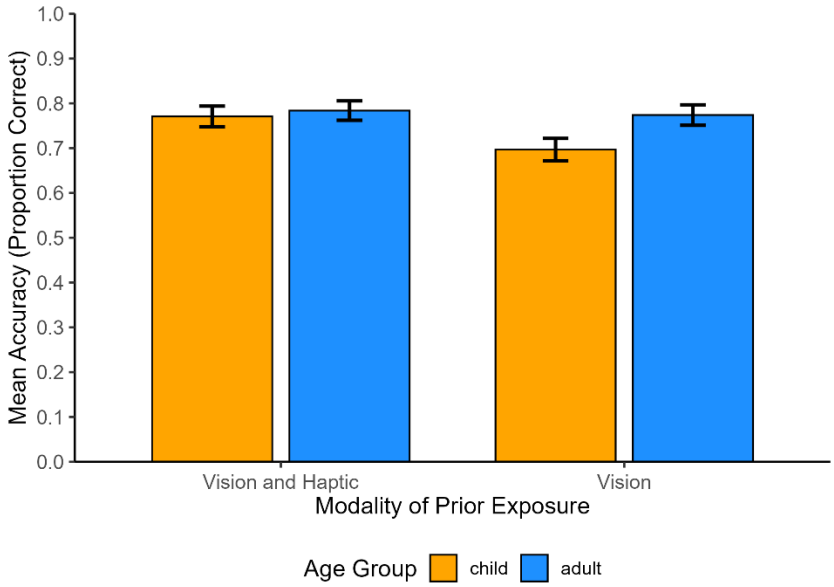
There was also a significant main effect of modality of prior exposure, $\chi^2(1) = 6.99$, $p = .008$, with accuracy in the combined visual and haptic experience ($M = 0.79$, $SD = 0.42$, 95% CI [0.76, 0.79]) higher than vision only experience ($M = 0.74$, $SD = 0.44$, 95% CI [0.72, 0.75]).

No effect of age was detected [$\chi^2(1) = 0.37$, $p = .543$]; however, the two-way interaction between modality of prior exposure and age group reached significance [$\chi^2(1) = 4.00$, $p = .05$] (see Figure 20). A post hoc pairwise comparison revealed no difference between children and adults performance in the combined visual and haptic prior exposure ($p > 0.5$); however, for participants in the vision-only group, children (M = 0.70, SD = 0.46;

95% CI [0.67, 0.72]) were significantly less accurate than adults (M = 0.77, SD = 0.42; 95% CI [0.75, 0.80]; $t(126) = -3.27$, p = 0.008).

Figure 20

Interaction effect between modality of prior exposure and age

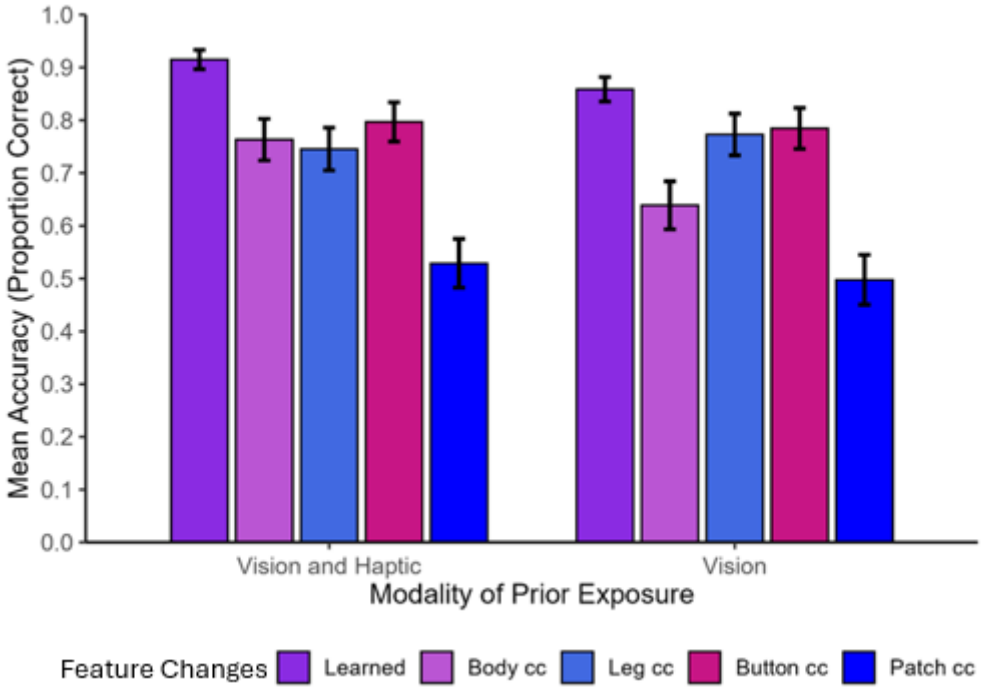


Note. Error Bars represent 95% confidence intervals

There was also a significant interaction between feature type and prior exposure modality [$\chi^2(4) = 30.58$, $p < 0.001$]. The participants in the combined vision and haptic prior exposure (M = 0.76, SD = 0.43; 95% CI [0.72, 0.80]) had significantly better accuracy performance than those in the visual condition (M = 0.76, SD = 0.43; 95% CI [0.72, 0.80]; $t(295) = 4.41$, p < 0.001) to objects with $body_{cc}$ feature change (see Figure 21 for details). The same pattern was observed for objects with leg_{cc} feature change: vision and haptic prior exposure (M = 0.75, SD = 0.44; 95% CI [0.71, 0.79]) compared to the visual condition (M = 0.77, SD = 0.42; 95% CI [0.73, 0.81]; $t(244) = 2.82$, p = 0.005). Finally, the two-way interaction between object type and age group as well as the three way interaction did not reach statistical significance (all *ps* > .12).

Figure 21

Interaction effect between modality of prior exposure and object type



Note. Error Bars represent 95% confidence intervals; Feature Changes refers to: learning objects i.e. no change, these are repeated objects from learning tasks, cc refers to objects with cross category features; the Vision only = body/leg colour, visuohaptic cross category features = button size/patch texture.

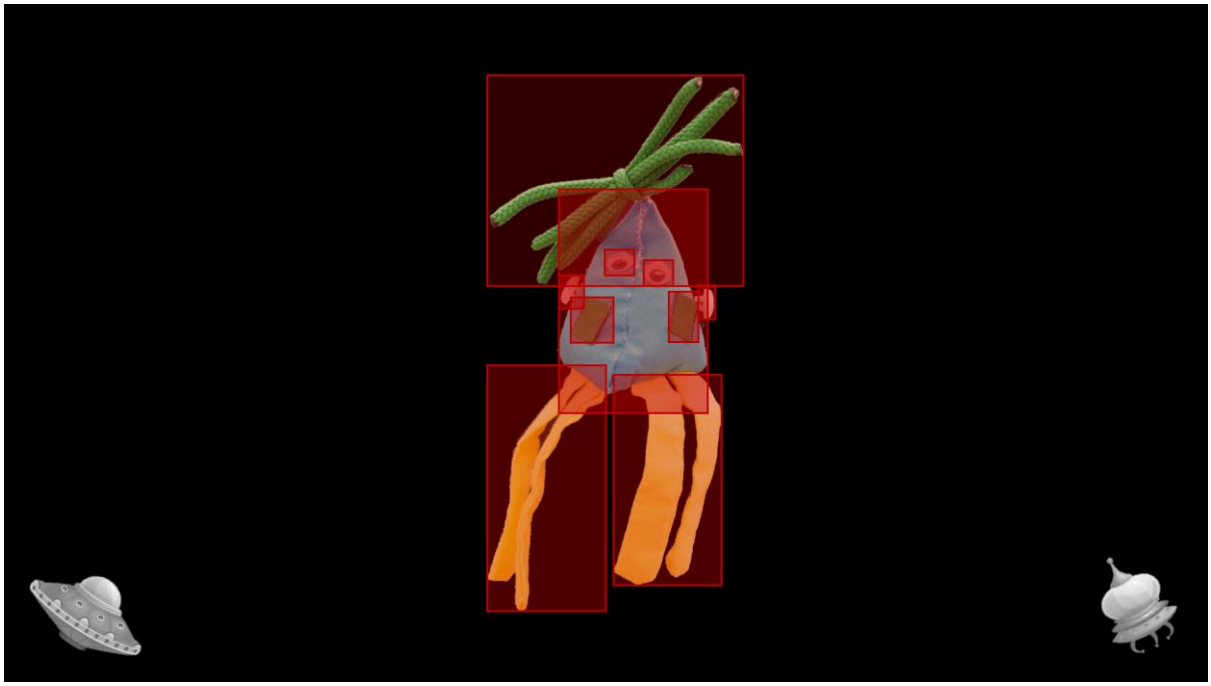
Learning Phase 1: Proportion of Fixations

To analyse eye movements, we focussed on the final block of learning trials across all participants (participants took different numbers of blocks to reach criterion). For this analysis, we isolated gaze points which fell beyond the bounds of our Area of Interest (AOI) regions (i.e. the surrounding screen), removing data which fell beyond the limits of the visual stimulus (17.19% of the data was removed as a consequence). We then characterised fixation thresholds as relative change of X,Y gaze coordinates which were below 30 degrees of visual angle per second velocity for a duration of at least 150ms. We isolated three different AOIs on the stimuli which encapsulated the full object: *visual features*

including body and legs; *visuohaptic features* including patches and button "ears"; as well as defining the *other object features*: eye regions and hair (see Figure 22 for illustration).

Figure 22

Defined AOI regions including visuohaptic, visual-only and other body features



Note. In the case of AOI regions which overlapped, if a fixation occurred within these boundaries they were first identified as occurring within both regions. Subsequent processing hierarchically assigned the fixation to the AOI type with the smallest area.

These AOI regions, due to the characteristics of the stimuli, varied in the area of the object they occupied. Therefore, in order to ensure that feature size was not a confound we first included AOI size as a covariate in our model (see Supplemental Materials 9 for proportion of fixations analysis & Supplemental 10 for duration of fixations analysis); however, feature area did not significantly affect the findings of the model and only the simplified model is reported here. We then calculated the number of fixations made within each AOI and used the proportion of fixations per participant across each of these regions for each trial as our measure in the following model. We conducted a linear mixed effects model using the lme4 package (Bates et al., 2015) to examine the proportion of fixations to

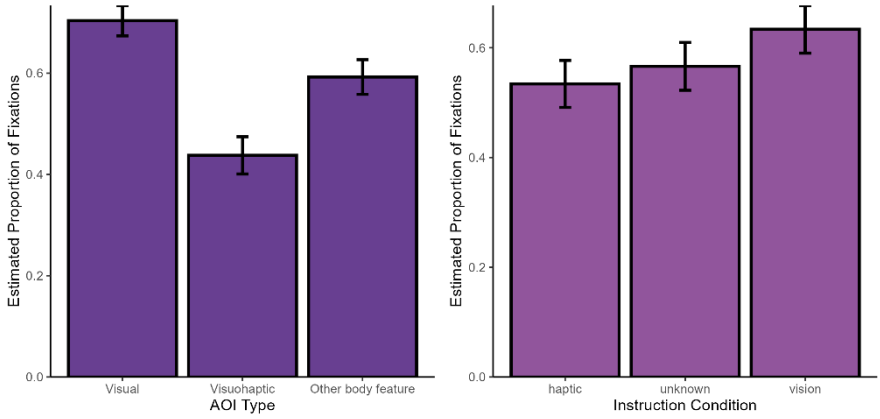
each AOI type as a function of age group (child versus adult), as well as across instruction conditions (visual test, haptic test, unknown testing condition). The model included random intercepts for participants and AOI types nested within participants to account for repeated measures. In total, the model included observations from 138 participants.

The model explained a moderate amount of variance, with a conditional $R^2 = 0.34$ and marginal $R^2 = 0.17$. There was a significant main effect of age group [$F(1, 86.21) = 4.67$, $p = .033$, $\eta_p^2 = .05$], indicating that adults and children differed overall in the proportion of fixations. Post-hoc contrasts, Bonferroni-adjusted averaged across conditions and AOI types, showed that adults ($M = 0.55$, $SD = 0.30$) allocated a significantly lower proportion of fixations to AOIs than children ($M = 0.61$, $SD = 0.29$; [$t(113) = -2.16$, $p = .033$]).

There was a significant effect of test instruction condition [$F(2, 86.17) = 5.42$, $p = .006$, $\eta_p^2 = .11$]. Pairwise comparisons showed that participants in the visual test condition ($M = 0.64$, $SD = 0.29$) allocated a lower proportion of fixations to the AOIs overall than those assigned to haptic test condition ($M = 0.53$, $SD = 0.30$; [$t(115) = -3.12$, $p = .005$, Bonferroni adjusted). There was no difference between the proportion of fixations made in the 'unknown' and vision ($p = .10$), nor between haptic and unknown ($p = .90$). There was also a main effect of AOI type [$F(2, 177.68) = 89.38$, $p < .001$, $\eta_p^2 = .50$]; the main effects are illustrated in Figure 23. Follow-up tests revealed that participants fixated significantly more on visual AOIs ($M = 0.\ 71$, $SD = 0.17$) than on visuohaptic AOIs ($M = 0.43$, $SD = 0.20$; [$t(185) = 13.33$, $p < .0001$]) and 'other' body features ($t(175) = 5.89$, $p < .0001$). Visuohaptic AOIs were also fixated more than 'other' body features ($t(229) = 7.16$, $p < .0001$). There was no evidence for two-way or three-way interactions between the factors (all $ps > .10$), suggesting that the effects of AOI type and condition on fixation proportions were broadly similar across age groups and conditions.

Figure 23

Mean Estimated Proportion of Fixations per AOI type and Instruction Condition



Note. Error Bars represent 95% confidence intervals. Proportion of Fixations is calculated using model estimates.

Learning Phase 1: Duration of Fixations

A linear mixed effects model was conducted using the lme4 package (Bates et al., 2015) to assess the potential differences in mean fixation duration across the AOI regions (visual-diagnostic, visuohaptic-diagnostic, other body features-distractors) for each age group (child versus adult) and their instruction condition (visual test, haptic test, unknown testing modality). A random intercept was included per participant to account for individual differences across the repeated measures. The model explained 33.7% of the variance, with a conditional R^2 of 0.37 and a marginal R^2 of 6.6%. The Type III ANOVA revealed a main effect of AOI type [$F(2, 212.70) = 7.88$, $p < .001$, $\eta_p{}^2 = .07$]. Pairwise comparisons (Tukey-adjusted) for AOI type indicated that fixations to visual AOIs (M = 0.351, SE = 0.011) were significantly longer than fixations to either the visuohaptic AOIs (M = 0.309, SE = 0.011; $t(231) = 3.24$, $p = .004$) or other body features (M = 0.308, SE = 0.011; $t(232) = 3.36$, $p = .003$). Fixations to visuohaptic AOIs and other body features did not significantly differ ($p = .996$). There was no effect of age group [$F(1, 113.12) = 0.03$, $p = .854$], or test instruction [$F(2, 113.14) = 0.62$, $p = .542$]. The interaction between age group and AOI type was not

significant [$F(2, 212.70) = 2.70$, $p = .07$, $\eta_p^2 = .02$], nor was the three-way interaction between age group, condition, and AOI type [$F(4, 212.69) = 2.18$, $p = .07$, $\eta_p^2 = .04$]. The other interactions were also not significant ($Fs < 0.62$, $ps > .39$).

Learning Phase 2: Proportion of Fixations

A linear mixed effects model was conducted using the lme4 package (Bates et al., 2015) to assess the potential effects of prior exposure to the objects in each modality (visual-only, or both visual and haptic), across children and adults on their proportion of fixations to defined AOI's. The model included fixed effects for all main effects and interactions, with random intercepts for participants and participant by AOI type to account for repeated measures within individuals across AOI types. The model was fitted using maximum likelihood estimation. The model explained a moderate proportion of variance in fixation proportions (*Conditional* $R^2 = .30$), with fixed effects alone accounting for 11% of the variance (*Marginal* $R^2 = .11$). The estimated variance for participant intercepts was 0.016 (SD = 0.126), for participant by AOI type intercepts the variance was 0.002 (SD = 0.040), and the residual variance was 0.062 (SD = 0.249).

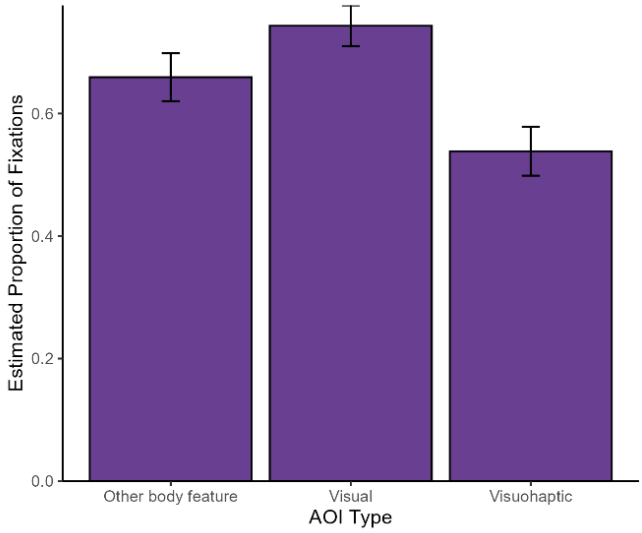
A Type III ANOVA with Satterthwaite's method revealed a significant main effect of Age Group [$F(1, 101.25) = 7.25$, $p = .008$, partial $\eta^2 = .07$] with a lower proportion of fixations made by adults ($M = 0.65$, $SD = 0.18$) than children ($M = 0.75$, $SD = 0.19$; [$t(122) = -2.64$, $p = 0.01$, $\eta_p^2 = 0.07$]) to a specific AOI region. The main effect of modality of previous exposure (visual vs. both visual and haptic exposure) was not significant [$F(1, 101.25) = 0.41$, $p = .522$, partial $\eta^2 < .01$]. However, there was a main effect of AOI Type [$F(2, 181.61) = 57.14$, $p < .001$, partial $\eta^2 = .39$]. A larger proportion of fixations were made to Visual AOIs ($M = 0.74$, $SD = 0.17$) were than either other body features ($M = 0.75$, $SD = 0.19$; [$t(182) = 4.30$, $p < 0.001$]) or visuohaptic AOIs ($M = 0.53$, $SD = 0.22$; [$t(186) =$

10.40, $p < 0.001$]). The proportion of fixations to other body features was more than the visuohaptic AOIs [$t(186) = 10.40$, $p < 0.001$] (see Figure 24 for more details).

No significant interactions were observed between the fixed factors of Age Group and prior exposure [$F(1, 101.25) = 0.53$, $p = .47$]; prior exposure and AOI Type [$F(2, 181.61) = 1.05$, $p = .35$]; or Age Group and AOI Type interaction [$F(2, 181.61) = 2.49$, $p = .09$, partial $\eta^2 = .03$]. The three-way interaction between our fixed effects was also not significant [$F(2, 181.61) = 0.53$, $p = .59$].

Figure 24

Mean Proportion of Fixations to areas of interest (AOIs) per trial averaged across participants



Note. Error Bars represent 95% confidence intervals, plotting estimated proportion of fixations calculated per trial per participant.

Learning Phase 2: Duration of Fixations

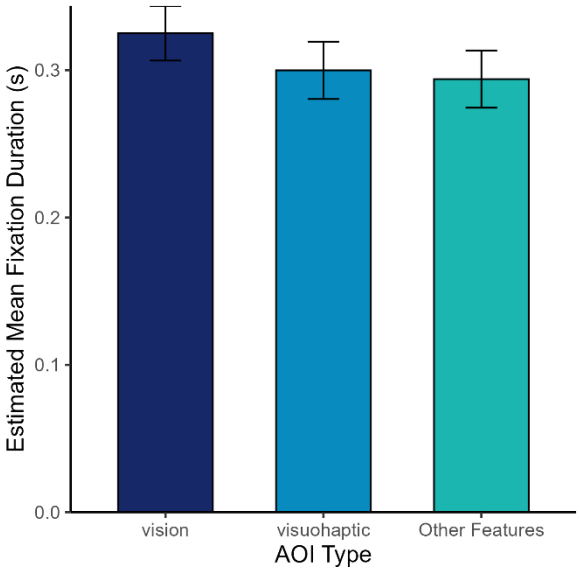
A linear mixed-effects model was fitted to examine the effects of age group (adult, child), modality of prior exposure (vision, haptic and vision), and AOI feature type (visual-only, visuohaptic, other features) on mean fixation duration. Participant was included as a random intercept to account for individual differences. The model showed good fit to the data, AIC = –582.53, BIC = –530.67, log-likelihood = 305.26. The conditional R^2 was .54

and the marginal R^2 was .11, indicating that most of the explained variance was attributable to between-participant variability.

The Type III ANOVA with Satterthwaite's approximation revealed a significant main effect of prior exposure [$F(1, 101.58) = 6.72$, $p = .011$, $\eta_p^2 = .06$], with post hoc pairwise comparisons (Tukey-adjusted) indicating that fixation durations were significantly longer for participants who had visual-only exposure ($M = 0.33$s, $SD = 0.11$, 95% CI [0.304, 0.347]) compared to those with both visual and haptic exposure ($M = 0.29$ s, $SD = 0.08$, 95% CI [0.265, 0.308], mean difference = 0.039 s, $SE = 0.015$ s, [$t(115) = 2.54$, $p = .012$]). There was also a main effect of AOI type [$F(2, 187.79) = 5.63$, $p = .004$, $\eta_p^2 = .06$] as shown in Figure 25, with post hoc comparisons showed that visual AOIs ($M = 0.33$ s, $SD = 0.10$, 95% CI [0.307, 0.344]) were fixated significantly longer than either visuohaptic AOIs ($M = 0.30$s, $SD = 0.12$, 95% CI [0.280, 0.319]; mean difference = 0.025 s, $SE = 0.010$ s, [$t(206) = 2.48$, p = .037]) or other features ($M = 0.29$s, $SD = 0.09$, 95% CI [0.275, 0.313]; mean difference = 0.031s, $SE = 0.01$s, [$t(206) = 3.07$, $p = .007$]). The difference in fixation duration between visuohaptic AOIs and other features was not significant ($p = .836$). The main effect of age group was not significant [$F(1, 101.58) = 2.74$, $p = .101$, $\eta_p^2 = .03$].

Figure 25

Mean Duration of Fixations compared across the defined AOI regions

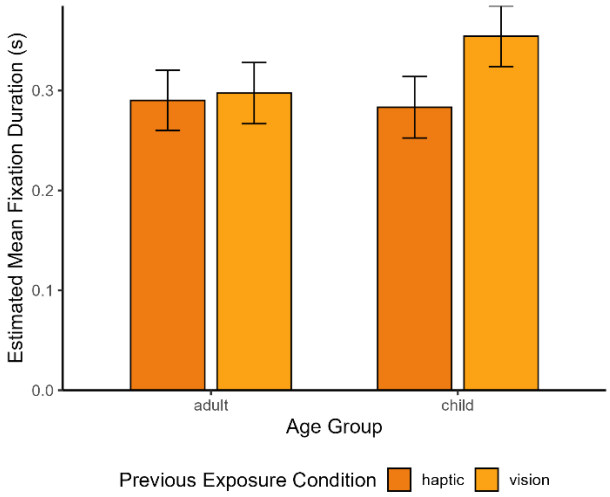


Note. Error Bars represent 95% confidence intervals.

There was a significant interaction between age group and prior exposure [F(1, 101.58) = 4.46, p = .04 , η_p^2 = .04]. Pairwise comparisons with Bonferroni adjustment revealed that children's fixation durations were significantly longer in the vision condition (M = 0.39s, SD = 0.13) compared to both adults in the haptic condition (M = 0.29s, SD = 0.07; mean difference = 0.06s, SE = 0.022s, [t(116) = –2.97, p = .022]) and children in the haptic condition (M = 0.29s, SD = 0.09; mean difference = 0.07s, SE = 0.022s, [t(119) = –3.25, p = .009]); see Figure 26 for details. There was no difference between adults and children in the vision condition, p = .061. None of the remaining interactions reached significance: age group by AOI type [F(2, 187.79) = 2.24, p = .109, η_p^2 = .02]; prior exposure by AOI type [F(2, 187.79) = 1.90, p = .15, η_p^2= .02]; and the three-way interaction was similarly non-significant [F(2, 187.79) = 0.87, p = .42, η_p^2 < .01].

Figure 26

Mean Fixation duration as compared across previous exposure conditions and age groups



Note. Error Bars represent 95% confidence intervals

Generalisation Test (Phase 2): Proportion of Fixations

In order to assess if the proportion of fixations across the defined AOI regions (visual, haptic, other body features) differed due to feature change type (object features with cross category visual (Vcc), visuohaptic (VHcc), or learned features), the age group (adult vs. child), and previous modality of exposure (vision only, visual and haptic), we conducted a linear mixed effects model. We first specified a maximal linear mixed-effects model using the lmer function from the lme4 package (Bates et al., 2015). The model included all fixed effects and their full interaction (age group × condition × stimulus type × AOI type), as well as random intercepts for participants and for each participant-by-within-subject factor combination (participant/ stimulus type and participant/AOI type). However, this model failed to converge and produced a degenerate Hessian with negative eigenvalues, as well as warnings about large eigenvalue ratios, indicating that the model was nearly unidentifiable. These convergence issues suggested that the random-effects structure was overparameterized relative to the available data.

Following best-practice recommendations for linear mixed-effects modelling (Barr et al., 2013; Matuschek et al., 2017), we simplified the model by retaining theoretically justified random intercepts for participants and for participant × within-subject factor combinations, while removing higher-order random slopes. The fixed-effects structure was also reduced to focus on the theoretically relevant two-way interactions between age group and each predictor, as well as the interaction between stimulus type and AOI type. This simplified model reflects the design of the study, in which age group and modality of prior exposure are between-subject factors, whereas stimulus type and AOI type vary within participants. Importantly, this model converged successfully without singular fit warnings and provided stable parameter estimates, while still accounting for the key sources of random variability in the data. To formally compare the fit of the maximal and simplified models, we conducted a likelihood ratio test using the anova() function. The simplified model (AIC = 865.0, BIC = 1042.3) provided a significantly better fit relative to model complexity than the original maximal model (AIC = 935.9, BIC = 1269.9), ΔAIC = 70.9, ΔBIC = 227.6. The likelihood ratio test indicated that the increase in model complexity in the maximal model did not result in a statistically significant improvement in model fit, $\chi^2(20) = 18.64$, $p = .54$.

The model fit indices indicated a conditional R^2 was .23, indicating that the model including both fixed and random effects explained 23% of the variance, whereas the marginal R^2 was .09, indicating that the fixed effects alone explained 9% of the variance. The intraclass correlation coefficient (ICC) was .15, reflecting substantial between-participant variability.

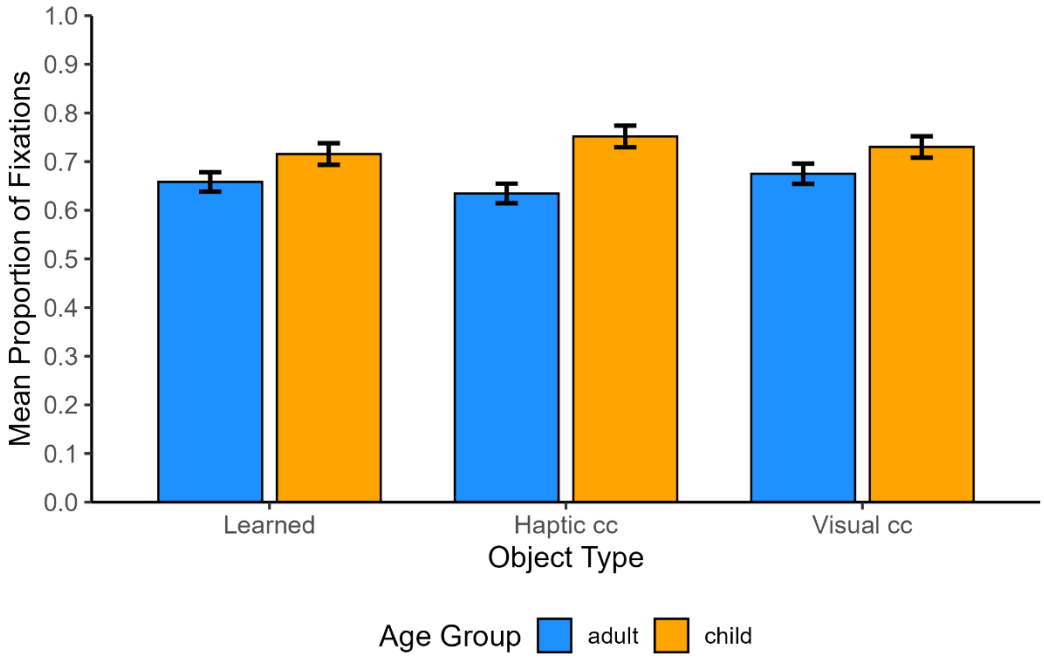
A Type III ANOVA (Satterthwaite's method) revealed significant main effects of age group [$F(1, 94.5) = 14.57$, $p < .001$, $\eta_p^2 = .13$], with a higher number of fixations by children ($M = .73$, $SD = 0.28$, 95% CI [.72, .75]) than adults ($M = .66$, $SD = 0.29$, 95% CI

[.64, .67]; $t(106) = 3.82$, $p = .0002$). There was also a main effect of AOI type [$F(2, 167.9) = 105.05$, $p < .001$, $\eta_p^2 = .56$]. For AOI type, participants directed significantly more fixations to visual features ($M = 0.77$, $SD = 0.27$, 95%CI [0.75, 0.78]) than other body features ($M = 0.66$, $SD = 0.30$, 95%CI [0.64, 0.68]; $t(172) = 7.77$, $p < .0001$) or visuohaptic diagnostic features ($M = 0.57$, $SD = 0.28$, 95%CI [0.56, 0.59]; $t(131) = 5.39$, $p < .0001$). A greater proportion of fixations were on other body than visuohaptic features ($t(209) = 3.20$, $p = .005$). The main effect of feature type did not reach significance [$F(2, 207.6) = 0.4$, $p = .67$, $\eta_p^2 = 0.003$].

There was a significant interaction between age and feature type [$F(2, 184.8) = 4.13$, $p = .018$, $\eta_p^2 = 0.04$], with post hoc pairwise comparisons indicating that adults had fewer fixations than children across all feature types (see Figure 27). Adults number of fixations was lower ($M = 0.64$, $SD = 0.29$, 95%CI [0.62, 0.66]) than children's ($M = 0.75$, $SD = 0.28$, 95%CI [0.73, 0.77]) for objects with a cross category Haptic feature (Haptic cc; $t(183) = -4.75$, $p < .0001$); also lower ($M = 0.68$, $SD = 0.29$, 95%CI [0.65, 0.70]) compared to children ($M = 0.73$, $SD = 0.28$, 95%CI [0.71, 0.75]) for objects with a cross category Visual feature (Visual cc; $t(187) = -2.50$, $p = .013$), and lower (adults ($M = 0.66$, $SD = 0.29$, 95%CI [0.64, 0.68]) compared to children ($M = 0.72$, $SD = 0.29$, 95%CI [0.69, 0.74; $t(178) = -2.57$, $p = .011$) for learned objects.

Figure 27

The mean proportion of fixations to each object modality exposure across age groups

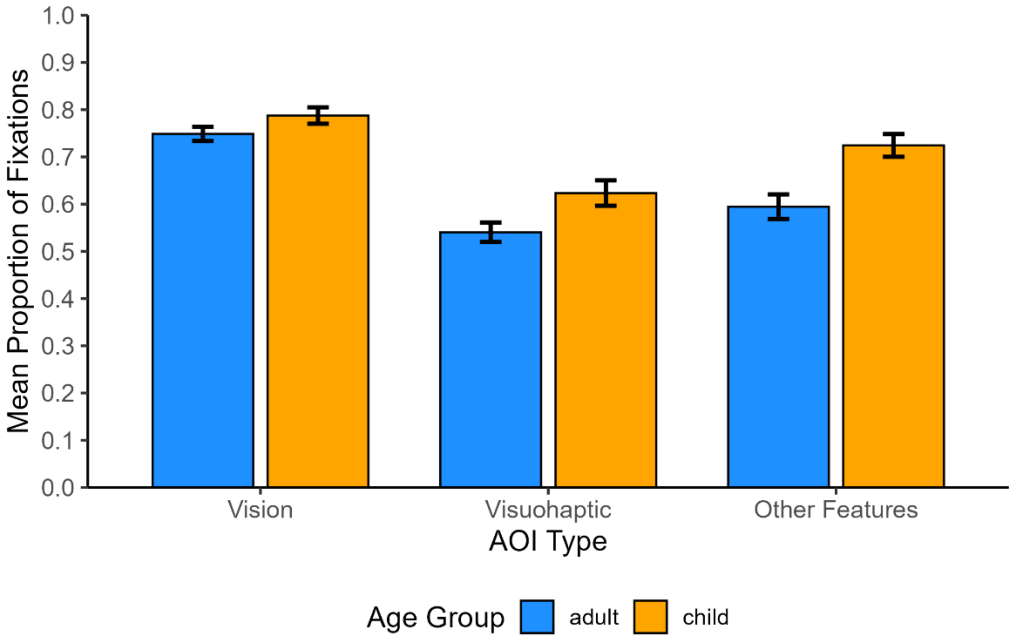


Note. Error bars represent 95% CIs; cc refers to cross category feature

There was a significant age group by AOI type interaction [$F(2, 167.8) = 5.87$, $p = .003$, $\eta_p^2 = .07$], as shown in Figure 28. The proportion of fixations by adults ($M = 0.60$, $SD = 0.30$, 95% CI [0.57, 0.62]) was significantly lower than children to other body features ($M = 0.72$, $SD = 0.28$, 95% CI [0.70, 0.75]; [$t(230) = -4.35$, $p < .0001$]) and visuohaptic AOIs (adult: $M = 0.54$, $SD = 0.28$, 95% CI [0.52, 0.56]; children: $M = 0.62$, $SD = 0.28$, 95% CI [0.60, 0.65]; [$t(221) = -3.34$, $p = .001$]). There was no difference between children and adults' proportion of fixations to visual AOIs ($p = 0.11$).

Figure 28

Mean proportion of fixations to each AOI region across age groups



Note. Error bars represent 95% CIs; Vision and Visuohaptic AOI types were diagnostic of category membership, other features refers to the distractor feature hair type and eye regions.

Generalisation Test (Phase 2): Mean duration of fixations

A series of linear mixed-effects models were fitted to examine fixation duration, varying in fixed- and random-effects structure, and compared using likelihood ratio tests and AIC and BIC values. The maximal model, which included the full four-way interaction between age group, condition, modality variant, and AOI type and random intercepts for participants and participant-by-factor combinations, resulted in a singular fit and poorer model performance (AIC = −3589.3; BIC = −3335.2). Subsequent models progressively simplified the random-effects structure. The minimal model, which included fixed effects of age group, object type, AOI type, and modality of previous exposure with their two-way interactions and random intercepts for participants only, provided the best overall fit to the data (AIC = −3623.6; BIC = −3509.2). Adding random intercepts for participant-by-AOI

type or participant-by-object variant did not significantly improve model fit [$\chi^2(1) = 0.00$, $p = 1.00$], and [$\chi^2(1) = 0.59$, $p = .44$], respectively. Similarly, adding higher-order interactions did not yield improvements in model fit relative to the minimal model [$\chi^2(3) = 0.43$, $p = .93$]. Marginal R^2 values were small across all models ($R^2_m = .019–.024$), indicating modest explanatory power of the fixed effects.

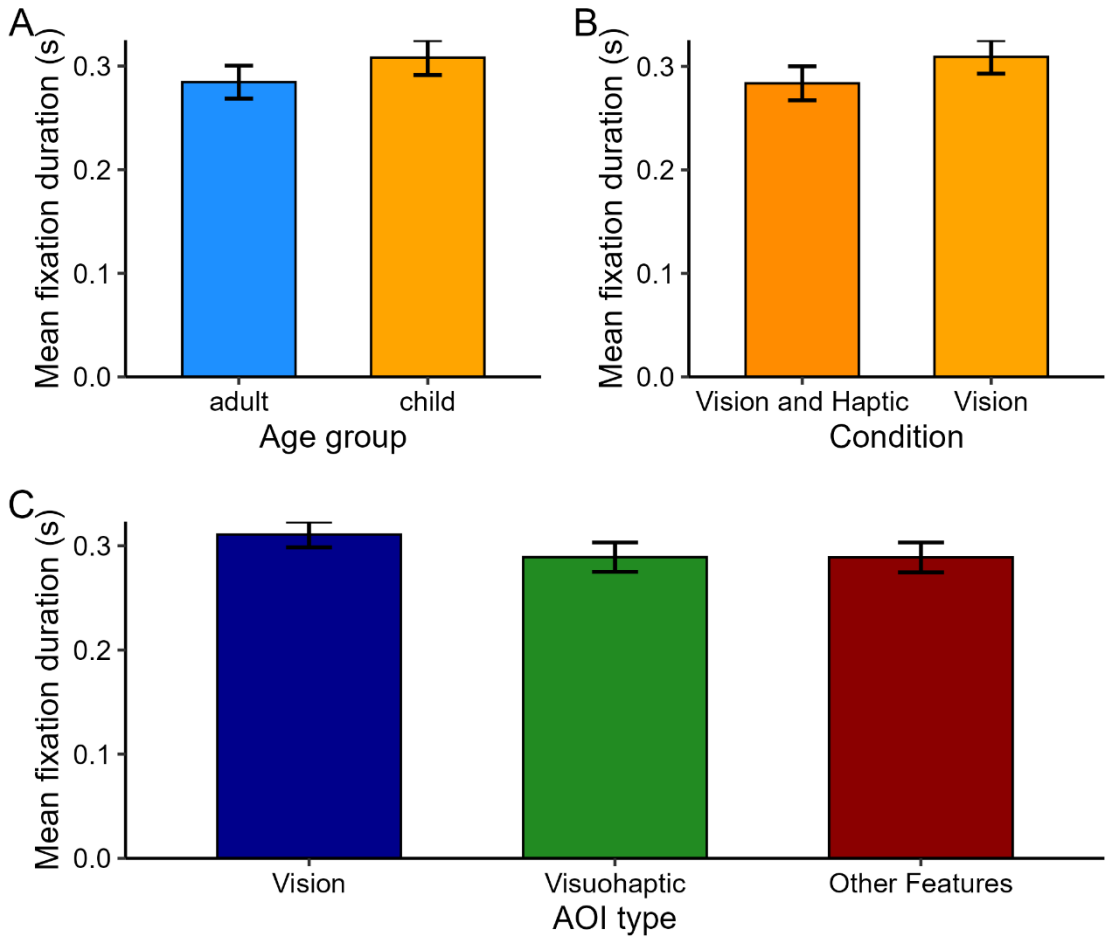
The model revealed a significant effect of age group [$F(1, 92.8) = 4.28$, $p = .041$, $\eta_p^2 = .04$], indicating that children ($M = 0.31$, $SD = 0.18$, 95% CI [0.30, 0.31]) had longer fixation durations than adults ($M = 0.29$, $SD = 0.15$, 95% CI [0.28, 0.30]). A Bonferroni-adjusted pairwise comparison confirmed this difference [$t(118) = −2.02$, $p = .046$]. A significant main effect of modality of previous exposure was also found [$F(1, 88.3) = 5.12$, $p = .026$, $\eta_p^2 = .05$]. Bonferroni-adjusted post hoc comparisons revealed that fixation durations were significantly shorter in the combined Vision and Haptic exposure condition (M = 0.29, $SD = 0.14$, 95% CI [0.28, 0.29]) than in the Vision-only condition ($M = 0.31$, $SD = 0.18$, 95% CI [0.30, 0.31]; $t(118) = −2.02$, $p = .05$). There was also a main effect of AOI type [$F(2, 4218.6) = 9.91$, $p < .001$, $\eta_p^2 = .005$]. Bonferroni-adjusted post hoc comparisons indicated that fixation durations were significantly longer to Visual-only AOIs ($M = 0.31$, $SD = 0.18$, 95% CI [0.30, 0.32]) compared to either Visuohaptic AOIs ($M = 0.28$, $SD = 0.14$, 95% CI [0.28, 0.29]; $t(153) = 3.57$, $p = .002$), or Other Features AOIs (M = 0.29, $SD = 0.14$, 95% CI [0.28, 0.30]; $t(164) = 3.54$, $p = .002$). There was no difference between Visuohaptic and Other Features AOIs ($p = 1.00$). All main effects are presented in Figure 29.

None of the two-way interactions between age group and object variant [$F(2, 4139.7) = 0.39$, $p = .68$], age group and AOI type [$F(2, 4218.4) = 1.06$, $p = .35$], age group and modality of previous exposure [$F(1, 88.3) = 2.96$, $p = .089$], and object variant and AOI

type [$F(4, 4141.6) = 0.27$, $p = .90$], were significant. Partial η^2 values for these effects were all $\leq .03$, indicating negligible effect sizes.

Figure 29

Mean duration of fixations across A) age groups B) Modality of prior exposure and C) AOI types



Note. Error bars represent 95% confidence intervals

Chapter 5

Discussion

The current research sought to examine how children and adults attend to stimuli during visual category learning and generalisation task when the modality of test is either known or unknown to be in the visual or haptic modality. In order to do this, we assessed how the manipulation of the expected testing modality may affect both gaze behaviour, category learning and subsequent categorisation accuracy in children compared to adults. We had hypothesised that adults would display more selective gaze behaviour, preferentially attending to modality relevant AOI regions, and that this would correspond to higher performance accuracy as compared to children.

When assessing category learning efficacy, we observed no significant difference between the number of trials required by adults compared to children. This was not in line with our hypothesis, nor with similar work in visual category learning which would suggest adults would outperform children (Huang-Pollock et al., 2011), however previous work has shown that for rule-based categories, children can perform as well as adults (Rabi & Minda, 2014). Where we did see a developmental effect was for those told their test would be in the visual modality, with adults acquiring the novel categories quicker than children, this same pattern was not observed for children and adults who were expecting a cross-modal haptic test or those in the unknown condition. However, there was a main effect of instruction condition on learning efficacy- whether the participant knew the modality they would be tested in, with those in the unknown condition learning in fewer trials. The unknown nature of their test modality and their increased uncertainty may have elicited more exploratory behaviour (Monosov, 2020; Sloutsky, 2016). These findings corresponded with a significant difference in both the proportion of fixations between children and adults (with children having a higher mean proportion), consistent with broader and less selective attention distributions (Blanco et al., 2019). There was also a main effect of instruction condition, with a lower mean proportion of fixations in the haptic

than the visual instruction condition. Similarly, AOI type led to differences in proportion of fixations with more fixations directed to the *visual* AOIs than either the *other body features* -the distractors, and the fewest overall directed to the *visuohaptic* AOIs. In the first test phase, though children had more fixations than adults in learning, adults were more accurate, particularly in the visual condition. Note, it was in the visual condition that the highest proportion of fixations was observed, and it was the vision-only features, body and leg colour, which received the highest proportion of fixations. Performance was significantly less accurate, though still well above chance, in the haptic test modality, with the lowest proportion of fixations during learning observed in this group and compared to visual AOIs shorter mean fixation durations were directed to visuohaptic AOI regions – the only accessible diagnostic features in the haptic test.

The second learning phase revealed that adults were significantly more efficient than children, as was expected. No further differences were observed, however, there was a trend for more efficient learning when participants had both visual and haptic prior exposure as observed in previous work (Kyler & James, 2024). Gaze behaviour in this task showed as in learning 1 that adults had a lower mean proportion of fixations than children, with the highest proportion of fixations again directed to visual AOIs followed by the other body features and then to visuohaptic AOIs. There were no differences observed between modality of prior exposure nor across instruction conditions which was contrary to our hypothesis. We predicted that modality of exposure would affect how people visually explore the stimuli, with greater attention directed to the features which you had access to during test, such as the visuohaptic AOIs for those in the haptic condition. In fact, the lack of observed changes between gaze behaviour in learning one as compared to learning two gives support to the argument that the gaze behaviour in this task was stimulus driven.

One interpretation of our eye tracking findings would be that the visual AOIs received a higher proportion of gaze simply due to the fact that they occupied a larger area. However, in an analysis of gaze behaviour to AOIs for both proportion of fixations (Supplemental 4) and duration of fixations (Supplemental 5) introducing size as a covariate to our model did not alter the effect of AOI type. The effect of AOI type remained even when controlling for the size of the AOIs themselves, thus perhaps some other difference in the feature saliency affected the proportion of gaze effects, as well as the duration of fixations being directed to the visual AOIs. Prior work on visual attention suggests that the phenomenon of preferential fixations can be driven by bottom-up salience for example body colour compared to the patch textural features (Nothdurft, 1993; Peters et al., 2005).

Nonetheless, it is when considering the cross-category feature manipulation that we get insight into which feature participants were predominantly reliant on to make their category judgements. If switching the feature to that of the opposing category leads to a drop in accuracy, it can be inferred that feature it is the determinant of their category judgements. As one would not have predicted when considering gaze behaviour, novel objects with a cross-category patch led to the lowest categorisation accuracy.

In the generalisation test there was also a main effect of modality of prior exposure with those who had both visual and haptic exposure performing with higher accuracy as compared to those with only visual exposure to the stimuli. This supports evidence that multisensory encoding strengthens object memory and category representations (Suzuki & Gyoba, 2008). It appears that visual-only exposure may have led to the incorporation of both body colour and patches as their category diagnostic features as opposed to the those who had visual and haptic prior exposure who appear to have relied on the object patches. Nevertheless, accuracy was significantly lower for all objects with cross category features compared to those that were repeated objects from learning. This may be due to the novel

nature of the stimuli in the final test phase, as opposed to incorporation of multiple features in participant category representations (Deng & Sloutsky, 2015). It may be interpreted from the behavioural findings that it was only the patch feature, or in the vision only modality of prior exposure, both body colour and patch which were the determinants of category judgements. Interestingly, it is here that we begin to see a true dissociation between gaze behaviour and categorisation, it has previously been said that the "eyes are the window to the representation" (Weichart et al., 2024), and that for adults during category learning allocation of attention shifts to the diagnostic, or most informative feature (Blair et al., 2009). However, in the final test, it was again *visuohaptic* AOIs which received the lowest proportion of fixations. The highest proportion was again directed to *visual* AOIs followed by the *other body features*. However, here we see an effect of age group with children directing significantly more fixations to the non-diagnostic *other body features* than adults. This suggests some evidence for more selective gaze behaviour in adults than children.

Taken together, these findings indicate that while adults and children can learn visual categories with comparable efficiency under some conditions, adults exhibit greater selectivity in both learning strategy and gaze behaviour once category rules are established. Manipulations of uncertainty and sensory modality at test influenced learning rates but had limited effects on gaze allocation, which remained largely stimulus-driven and biased toward visually salient features. The combined behavioural and eye-tracking data therefore suggest partial independence between visual attention and the decision processes underlying category learning. More broadly, the results contribute to developmental and multisensory learning research by showing that visual–haptic exposure enhances generalisation accuracy, yet attentional selectivity to diagnostic features may depend more on perceptual salience than on modality relevance.

Chapter 6

General Discussion

Chapter 6

Introduction

In the first chapter of this thesis a review of the literature concerning the development of unisensory and multisensory processing from birth across development into adulthood was outlined. Furthermore, how this developmental progression affects the recognition and categorisation of objects as well as the processes underlying the formation of novel object categories was examined. The emergence of multisensory and cross-modal benefits as compared to unimodal exposure for behavioural performance was discussed with reference to both to object recognition as well as categorisation. The vast majority of research in the areas of multisensory perception have been conducted in infant populations or in adults, with the study of multisensory object category learning in childhood remaining particularly understudied to this point. To address the question of how children aged 4-13 years utilise multisensory information to form object categories a series of experiments were conducted. These experiments incorporated combinations of visual, auditory, and haptic information and through a series of category learning and categorisation tasks we wished to assess the relative influence of each modality as well as any potential cross-modal or multisensory behavioural benefits to category acquisition. Furthermore, a specific focus was directed to examining the nature of the category representations formed in memory upon successful category acquisition.

We assessed categorisation performance to familiar object using audiovisual object and visual contextual information (Chapter 2), as well as the acquisition of novel categories using audiovisual dynamic objects in children aged 5-13 years (Chapter 3). We further compared visuohaptic as opposed to unisensory (vision-only, haptic-only) category learning and subsequent cross-modal categorisation and generalisation performance in children aged 4-13 years (Chapter 4). Finally, in order to assess the top-down influence of selective attention we utilised eye-tracking measures in a visual category learning task of novel visuohaptic objects in children and adults to assess the efficacy of learning when one

is informed of their future modality at test (vision or haptic; Chapter 5). This discussion will contextualise the research findings presented in this thesis with theories of category learning, categorisation and multisensory development as well as deliberate on the potential implications this work has for future investigations.

Summary of Main Findings

Chapter 2 investigated 5–13-year-olds categorisation of familiar objects based on multisensory and contextual cues. Our results indicate a general enhancement in categorisation performance with age, reflected in both higher accuracy and faster reaction times. This developmental trend aligns with previous findings (Bjorklund, 1985; Broadbent & White et al., 2018). Additionally, we observed a clear effect of modality: participants were less accurate and slower when categorising auditory-only object cues compared to visual-only or audiovisual cues. Contrary to prior evidence suggesting a multisensory benefit (Broadbent et al., 2020b), we did not observe any performance advantage for audiovisual over visual-only cues in either categorisation accuracy or response times.

In Chapter 3 an investigation of audiovisual category learning was conducted where we utilised novel moving objects with correlated sounds to assess the relative contribution of visual shape, movement and sound to the formation of novel category representations in children and adults. We observed no difference in category acquisition efficacy; however, when testing the nature of the representations acquired, both children and adults displayed evidence of associative use or incorporation of both motion and sound into their representations in memory, and adults were significantly more accurate than children.

In Chapter 4 we assessed category learning in either visual-only, haptic-only or visuohaptic modalities across age groups ranging from 4-13 years. Firstly, 4–6-year-olds had difficulty acquiring these categories, however, across learning conditions, haptic and visuohaptic features appeared to be of most use during categorisation. Older children were

better at categorisation test, however, all children from 4-13 years could generalise their learning to novel objects and for visual and haptic learners, doing so in the new visuohaptic context.

Chapter 5 investigated the top-down influence of selective attention on category learning, we assessed the gaze behaviour of children and adults in a visual category learning task when participants anticipated either a visual, haptic, or unknown (visual or haptic) test modality. We found no clear differences in gaze behaviour between our test instruction conditions, however, children fixated upon our stimuli both with more frequency as well as for longer durations than adults. There was no clear evidence of the strategic allocation of attention to diagnostic cues of use for subsequent test. Nonetheless, we found that the visuohaptic texture cue was the greatest determinant of categorisation judgements. Behaviourally, adults outperformed children in the visual test condition, and in later experiment phases, having exposure to the stimuli both visually and haptically, led to higher performance accuracy than visual-alone.

Implications of the findings and insights into outstanding issues in the field

Does children's categorisation performance benefit from object specific audiovisual cues or visual contextual cues? Does this change with age?

The findings of Chapter 2 contribute to understanding how multisensory and contextual information interact across cognitive development during the categorisation of familiar objects. Overall, results indicate a developmental progression between five and thirteen years of age, characterised by improved accuracy and faster responses with increasing age. This pattern aligns with theories of progressive conceptual refinement and increasing information-processing efficiency (Bjorklund, 1985; Siegler, 1996), suggesting that as children mature, they become more adept at extracting and integrating perceptual and conceptual information. From a neurocognitive perspective, these improvements likely

reflect continuing maturation of the ventral visual stream and frontoparietal control systems that support object recognition and attentional regulation (Casey et al., 2019).

Consistent with previous work, children showed clear modality-specific asymmetries, performing best with visual-only and worst with auditory-only cues. This behavioural profile supports evidence that visual object processing attains functional maturity earlier than auditory processing (Robinson & Sloutsky, 2004a; Litovsky, 2015). The lack of age-related improvement in auditory-only categorisation speed further suggests a developmental constraint in auditory object recognition, consistent with asynchronous sensory maturation models, in which auditory cortical and temporal integration mechanisms develop more slowly than visual systems (Hensch, 2005; Moore, 2012; Krishnan & Gandour, 2013). Such findings highlight the distinct developmental trajectories of the auditory and visual modalities and their implications for how children combine sensory information during categorisation. This behavioural pattern may also be explained by the modality appropriateness framework (Welch & Warren, 1980; Sloutsky & Napolitano, 2003), which proposes that perceptual weighting is biased toward the most reliable modality. The absence of a measurable advantage for audiovisual cues over visual-only conditions aligns with the principle of inverse effectiveness (Stein & Meredith, 1993), whereby multisensory gains emerge primarily under perceptual uncertainty or degraded unisensory input. According to Multisensory Integration Theory (Wallace & Stein, 2007; Nardini et al., 2008), integration depends on the relative reliability of each cue; thus, when visual categorisation is near ceiling, additional auditory input provides limited benefit. Empirical work similarly indicates that multisensory integration continues to mature through late childhood and is most effective when stimuli are ambiguous or noisy (Gori et al., 2008; Hillock et al., 2011).

The finding that semantic scene context enhanced performance only for auditory conditions provides behavioural support to the semantic facilitation account of crossmodal perception. When perceptual input is less informative, as in auditory-only conditions, congruent scene context may exert a top-down influence that enhances recognition and categorisation (Bar, 2004; Davenport & Potter, 2004; Shafiro et al., 2012). Our observed behavioural results when considered within a predictive coding framework (Clark, 2013), may reflect the use of contextual expectations to constrain sensory inference. In contrast, when visual cues dominate, additional contextual information may add little benefit or impose attentional competition (Porcu et al., 2014), which may be reflected in our behavioural finding of slower audiovisual categorisation in the presence of scene backgrounds. Perhaps indicating developmental constraints in attentional resource allocation (Sloutsky & Robinson, 2008; Goldenberg & Johnson, 2015), where concurrent processing of object, scene, and crossmodal information exceeds children's cognitive capacity. The mixed pattern of scene congruency effects may also relate to the level of categorical decision making. Previous research indicates that congruent contexts can promote superordinate categorisations (e.g., "wild animal"), whereas incongruent scenes may facilitate basic-level labels (e.g., "lion") (Murphy & Wisniewski, 1989). These findings are consistent with object–scene interaction models (Biederman et al., 1982), suggesting hierarchical category processing modulated by contextual coherence.

Finally, the broader implications of these results suggest that children's categorisation of familiar objects may depend primarily on established semantic representations, reducing reliance on multisensory or contextual cues (Mandler, 2000; Oakes & Rakison, 2003). This contrasts with evidence that such cues are more influential in novel category formation (Broadbent et al., 2020b). Overall, the findings support the view that multisensory and contextual integration in categorisation is developmentally

dynamic and context-dependent, with the greatest benefits emerging when perceptual uncertainty or conceptual novelty is high.

Does audiovisual category learning efficacy vary across age groups, do object category representations incorporate motion and sound and does this vary by age?

The findings of Chapter 3 advance our understanding of how the availability of multisensory information supports category learning and categorisation across development. Although children and adults achieved comparable learning outcomes, the behavioural findings at test may be considered consistent with the formation of cross-modal associative representations rather than with obligatory multisensory integration. The enhanced performance in adults when compared to the developmental group as a whole however, could support interpretations that there is a transition from a cross-modal associative representation in childhood to an integrated multisensory representation in adulthood. This interpretation, though speculative and in need of future investigation, provides support for multisensory object representation theories (Murray, Lewkowicz et al., 2016) and embodied cognition accounts proposing that perception and action jointly shape conceptual knowledge (Barsalou, 2008). Specifically, the selective benefit for categorisation when object motion and sound were presented in the same combination as during learning suggests that participants encoded the co-occurrence of shape, motion, and sound features, and that successful categorisation was facilitated when this learned configuration was reinstated. This interpretation aligns with developmental accounts emphasising experience-dependent associative learning across modalities which support multisensory development (Lewkowicz & Kraebel, 2004; Lewkowicz, 2012). The absence of strong age-related differences in our 5–13-year-old participants in how motion and sound influenced categorisation suggests that the ability to form and retrieve cross-modal associations may be relatively stable across middle childhood. This pattern is consistent

with proposals that children are highly sensitive to statistical regularities and co-occurrences across sensory inputs (Kirkham et al., 2002; Ecker et al., 2013), and that these mechanisms can support learning even when optimal multisensory integration has not yet fully matured. The ability of both children and adults to generalise categorisation when motion or sound varied indicates that their category representations are robust yet adaptable, consistent with predictive processing models positing that perception involves probabilistic inference under uncertainty (Clark, 2013). Nevertheless. performance did decline when either motion or sound deviated from the learned pairing, which potentially supports that feature conjunctions, rather than individual cues, were encoded during learning. Crucially, this effect cannot be fully explained by familiarity with individual features, as both motion types and sounds were experienced across conditions. Instead, the results point toward sensitivity to specific audiovisual contingencies, consistent with frameworks that emphasise temporal alignment and contingency as key drivers of cross-modal learning (Bahrick & Lickliter, 2000; Ecker et al., 2013). From this perspective, the present findings provide evidence that both children and adults utilise structured regularities in multisensory input to build category representations that incorporate visually dynamic and auditory information.

Taken together, the findings suggest that both children and adults form multifeatural category representations that incorporate shape, motion, and sound. Furthermore, developmental differences between adults and children may reflect changes in the efficiency and robustness of multisensory processing.

Do children benefit from multisensory (vh) or unisensory (v/h) category learning and how does this affect categorisation performance?

In Chapter 4 we sought to examine potential developmental differences in the processes of visuohaptic object category learning, categorisation, and generalisation.

Children between four and thirteen years successfully formed novel categories in both unisensory and visuohaptic conditions, however, there was no evidence that visuohaptic learning enhanced acquisition relative to vision or haptics alone. This absence of a multisensory advantage suggests that multisensory exposure does not necessarily facilitate category learning when unisensory cues provide reliable information or when multisensory processing imposes greater cognitive demands. The behavioural performance observed here could be interpreted using the principle of inverse effectiveness (Stein & Meredith, 1993), which predicts that multisensory gains emerge primarily when individual sensory inputs are weak or ambiguous. They could similarly be considered as behavioural evidence which aligns with Bayesian cue-reliability models (Ernst & Bülthoff, 2004), indicating that the efficiency of cue integration is context-dependent and constrained by task complexity. Although younger children were more likely to fail to reach learning criterion, successful learners across age groups displayed comparable learning rates, suggesting that by early school age the core mechanisms for category acquisition are established, and the subsequent improvements may be largely reflective of the maturation of attentional and executive processes (Bjorklund, 1985; Siegler, 1996).

At test, categorisation performance improved with age, reflecting developmental enhancement in the ability to generalise category knowledge to novel exemplars. This age-related gain may stem from progressive refinement of multisensory exposure via the proposed process of crossmodal calibration, by which one sense (e.g., touch) calibrates another (e.g., vision) to achieve greater perceptual accuracy (Gori et al., 2010; Gori, Sandini et al., 2012; Gori, 2015). Enhanced calibration and increasing working-memory capacity (Casey et al., 2019) may allow older children to form more abstract and transferable category representations, integrating multisensory cues more efficiently. The youngest group's reduced generalisation performance and marginally higher trial requirements in the

visuohaptic learning condition is potentially suggestive of greater difficulty processing redundant or competing features, consistent with feature-selection theories that emphasise developmental improvements in focusing on diagnostic over non-diagnostic information (Sloutsky, 2010).

To identify which features supported categorisation, we examined the effects of cross-category feature changes during a generalisation test. Alterations to haptic and visuohaptic properties disrupted performance more than changes to visual features, suggesting that participants placed greater functional weight on haptic information under these task conditions, rather than relying exclusively on vision (Ernst & Banks, 2002; Welch & Warren, 1980). This pattern may reflect selective attention during learning, whereby children prioritise features that are more salient or readily encoded (Sloutsky, 2010; Sloutsky & Fisher, 2004). Active haptic exploration may have enhanced the distinctiveness of object properties, biasing their use in subsequent categorisation, while visual information remained part of the representation but exerted less influence on behavioural responses. Notably, the current design cannot disentangle whether this effect reflects the informational value of haptic input or attentional allocation across modalities. In visuohaptic conditions, children may have encoded multisensory feature conjunctions (Ernst & Bülthoff, 2004). This distinction is particularly relevant given the visuohaptic nature of the generalisation test, where both visual and haptic signals were simultaneously available. It is plausible that children encoded combined feature representations, drawing on both modalities, but that changes to the haptic component had a larger behavioural impact, for visuohaptic and haptic learners, because it carried greater weight within the learned representation. Consequently, the disruption observed following changes to visuohaptic features should not be taken as evidence that categorisation was based purely

on touch, but rather that multisensory representations may have been asymmetrically weighted.

This interpretation is consistent with reliability-weighted cue combination accounts (Ernst & Banks, 2002), as well as developmental evidence that children do not always integrate sensory cues optimally and may rely on a single modality depending on task demands (Gori et al., 2008; Burr & Gori, 2012; Nardini et al., 2008). From a developmental perspective, these findings align with interactive specialisation frameworks, which propose increasing efficiency and coordination across systems over time rather than fixed modality dominance (Johnson, 2011). Limitations in attentional control and working memory may further constrain children's ability to process multiple cues simultaneously, encouraging selective reliance on the most accessible information (Casey et al., 2005). Crossmodal calibration accounts additionally suggest that haptic input may provide a stable reference for learning object structure in certain contexts (Gori et al., 2010, 2012). Taken together, the findings indicate that children form multisensory category representations that incorporate both visual and haptic information, but that categorisation behaviour may reflect context-sensitive cue weighting shaped by attention, reliability, and developmental constraints, rather than fixed modality dominance. This outcome underscores that perceptual systems underpinning conceptual organisation are calibrated through multisensory experience and that haptic processing provides a crucial scaffold for the development of stable and generalisable category knowledge.

What object features (v\|vh) receive increased attention from children and adults in visual category learning when one is informed of the testing (v/h) modality?

In Chapter 5 we investigated how children and adults allocate attention during a visual category learning and generalisation task, more specifically, we wished to compare how expectations regarding upcoming test modality, whether un/known to be visual or

haptic, influences gaze behaviour by guiding participants' visual attention toward modality-relevant features, thereby enhancing learning efficiency, and subsequent categorisation accuracy. Adults and children achieved comparable efficiency in visual category learning overall, but adults learned faster when the upcoming test was explicitly visual, whereas this advantage disappeared under cross-modal or unknown test expectations; together with the finding that participants in the unknown condition learned in fewer trials, these results imply that the core mechanisms for category acquisition are present and capable by middle childhood and that uncertainty about task demands promotes more exploratory sampling that can improve learning efficiency. The latter aligns with neural and behavioural work showing that uncertainty engages information-seeking systems and systematic exploration in both adults and children (Monosov, 2020; Blanco et al., 2021).

Gaze data further indicate developmental refinement in attentional selectivity: children produced more, but less selective, fixations (including to non-diagnostic "other body features"), whereas adults produced fewer and shorter fixations potentially reflecting higher efficiency once category rules were acquired. This is consistent with eye-tracking evidence that adults increasingly allocate gaze to diagnostic features during category learning and with developmental work showing broader, less selective attention in childhood (Blair et al., 2009; Blanco & Sloutsky, 2022). However, where our findings diverge from prior evidence is that we observed gaze to be allocated in a largely stimulus-driven manner and biased toward visual AOIs across learning phases and instruction conditions, even after controlling for AOI size, implicating bottom-up salience rather than prospective modality relevance as the principal driver of fixation selection; this observed behaviour may be partially interpreted through the lens of classic and computational work demonstrating that feature contrast and saliency-map computations strongly guide fixations in complex scenes (Nothdurft, 2002; Peters et al., 2005; Itti & Koch, 2000). The

dissociation observed between fixation patterns and the features that actually determined category decisions (poorest accuracy when the cross-category patch changed, despite low fixation to visuohaptic AOIs) indicates partial independence between overt attention and the internal representations that guide classification, tempering strong "eye-mind" assumptions in category learning and echoing broader demonstrations that decision-relevant feature weighting need not be mirrored directly in overt gaze.

Together, these findings suggest that attentional control and decision processes are coupled but separable: fixations can reflect salience-driven sampling while categorisation relies on feature-level evidence accumulated from prior exposure. The generalisation results sharpen this interpretation: participants with visual and haptic prior exposure outperformed those with visual-only exposure, indicating that multisensory encoding strengthens object memory and yields more robust category representations even when subsequent access is visual; recent empirical work similarly shows facilitation of recognition following multisensory (visuo-haptic or visuo-motor-haptic) experience (Kyler & James, 2024; Lacey et al., 2014). Nevertheless, haptic test performance was less accurate and visuohaptic AOIs received the fewest fixations, suggesting that accessibility and salience constraints—not simply expected modality—shape attentional allocation; this fits perceptual-load and salience accounts in which capacity limits favour dominant visual features when multiple diagnostic sources compete (Peters et al., 2005; Itti & Koch, 2000). Finally, the improved adult accuracy in the visual condition and children's greater fixations to non-diagnostic features observed in the final test phase indicate a developmental shift from diffuse, stimulus-bound exploration toward selective weighting of informative cues, consistent with models in which executive maturation sharpens selective attention and supports more efficient hypothesis-testing during category learning (Blair et al., 2009; Blanco & Sloutsky, 2022). In sum, these findings support a developmental account in which

uncertainty-driven exploration, salience-biased gaze, and multisensory encoding jointly shape category learning: children and adults can learn at similar rates under some conditions, but adults show greater strategic selectivity; gaze remains predominantly bottom-up; and visuo-haptic exposure enhances the generalisation of category knowledge by enriching underlying representations.

Developmentally Mediated Changes in Bottom-Up Processing Guiding Category Formation, Categorisation, and Generalisation

Across Chapters 2–5, the current research extends understanding of how the bottom-up processing of sensory information evolves through childhood to support object category learning and categorisation from multisensory input. The findings suggest that sensory-driven cross-modal association of cues, such as vision, audition, and haptics, provides the perceptual foundation from which more abstract multisensory object representations are constructed. The relative weighting of these cues shifts with both neural maturation and experiential calibration, reflecting an adaptive system that balances crossmodal reliability and temporal congruence to support stable category representations.

The developmental asymmetries observed between auditory, visual, and haptic conditions (Chapters 2 and 4) provide support to asynchronous sensory maturation models, whereby vision achieves functional precision earlier than auditory and tactile systems (Hensch, 2005; Krishnan et al., 2013; Litovsky, 2015). Younger children's weaker auditory-only categorisation (Chapter 2), combined with the suggestion of potential haptic dominance in visuohaptic tasks (Chapter 4; Broadbent et al., 2020; Holmes et al., 2024), supports the argument that the reliability of unisensory input constrains early multisensory categorisation. These findings when considered alongside those in Chapter 3, where children from age five associated dynamic audiovisual object features into their category representations lend support to the dual-trajectory hypothesis (Rohlf et al., 2020; Bruns &

Röder, 2023), proposing that multisensory integration and crossmodal calibration develop along partially independent timelines. Integration may play a stabilising role to sensory mappings early, whereas recalibration refines perceptual precision across modalities in later childhood.

The absence of a consistent advantage for visuohaptic learning (Chapter 4) when interpreted with consideration to both the principle of inverse effectiveness (Stein & Meredith, 1993) and Bayesian cue-reliability frameworks (Ernst & Bülthoff, 2004), is consistent with the position of Newell et al., (2023) that multisensory facilitation is context-dependent rather than ubiquitous. In Chapter 2, semantic scene context facilitated categorisation for auditory-only object cues whereas there was no benefit when the audiovisual object cues were paired with that semantic scene context. Furthermore, in Chapter 5, cross-modal exposure enhanced category generalisation relative to vision-alone. This benefit supports a conditional facilitation model, in which the perceptual system dynamically adjusts the contribution of each modality according to its relative precision and task relevance (Gori, 2015; Nardini et al., 2016).

Findings from Chapter 5 further support the dominant role of bottom-up salience to gaze allocation throughout development. Both children and adults fixated predominantly on visual features irrespective of test-modality expectations, even after controlling for AOI size. This supports saliency-map models of visual attention (Itti & Koch, 2000; Nothdurft, 2002; Peters et al., 2005), suggesting that perceptual contrast continues to guide fixations automatically across development. However, the narrowing of attentional breadth with age, reflected in children's higher proportion of fixations to non-diagnostic features during the final test phase, indicates that in adulthood bottom-up capture becomes increasingly constrained by emerging top-down control (Casey et al., 2005; Sloutsky, 2010). Collectively, these findings support the view that bottom-up mechanisms provide the

foundational input for category learning and generalisation, which become more efficient and selective with development as calibration precision and crossmodal weighting improve.

Developmentally Mediated Changes in Top-Down Processing Guiding Category Formation, Categorisation, and Generalisation

This body of research further provides insight into the varying influence of top-down predictive and attentional mechanisms progressively shape multisensory category learning and generalisation across development. The behavioural evidence, when considered with reference to the predictive-coding framework in which the brain constructs hierarchical generative models that anticipate sensory input (Clark, 2013; Friston et al., 2015), suggests that the developmental improvements in categorization are driven by accumulated perceptual experience and a progressively refined ability to use expectations (Goldstone, 1998; Kersten et al., 2004), contextual cues (Medin & Shaffer, 1978), and uncertainty (Griffiths et al., 2010) to guide category learning.

The finding that participants in the unknown test-modality condition learned fastest (Chapter 5) suggests that uncertainty can enhance exploratory patterns in category learning tasks, which could be behaviourally consistent with computational models of active inference (Friston, 2010) with prior evidence that uncertainty engages prefrontal–dopaminergic systems driving adaptive learning (Monosov, 2020; Bonawitz et al., 2023). This extends prior work by indicating that the strategic adaptation of exploration, rather than prediction alone, constitutes a potential top-down mechanism in developmental multisensory categorisation. The gaze findings in Chapter 5 further extend theories of attentional weighting and gaze–decision coupling. Adults produced fewer, more selective fixations, whereas children showed broader, less efficient visual sampling. Crucially, categorisation accuracy depended on diagnostic visuohaptic features that received

relatively few fixations in both age groups, indicating a partial dissociation between overt attention and representational weighting, a phenomenon increasingly recognised in eye-tracking and decision-modelling research (Gao et al., 2024; Weichart et al., 2024). Finally, the advantage of visual–haptic crossmodal exposure for generalisation (Chapter 5) aligns with the view that exposure to mutually redundant multisensory object features enhances predictive stability, leading to more robust and transferable category representations (Clark, 2013; Murray & Wallace, 2016). Developmentally, this suggests that improvements in top-down prediction and error correction, rather than sensory integration alone, underpin children's increasing capacity to generalise.

Together, these findings refine current theory by demonstrating that the interaction between bottom-up sensory input and top-down attentional control is not additive but developmentally dynamic. Consistent with predictive learning accounts of multisensory categorisation (Newell, 2023), the results reveal that bottom-up reliability constrains early integration, whereas top-down expectation, selective attention and exploratory patterns progressively enhance efficiency and flexibility. In sum, this thesis advances current multisensory categorisation theory by demonstrating that while top-down predictive learning mechanisms become increasingly influential with development, they remain fundamentally dependent on the precision of underlying sensory processing. In other words, mature category learning reflects a balance between accurate perceptual encoding and the efficient use of predictive models to guide interpretation and generalisation.

Developmental Dynamics of Multisensory Object Category Learning

The present thesis advances current theories of multisensory category learning in children by proposing a developmentally dynamic framework in which category formation, categorisation, and generalisation emerge from the interaction between bottom-up sensory information and progressively refined top-down control processes. Across Chapters 3–5,

the findings converge to suggest that multisensory category learning is not solely determined by sensory integration, but by how learners select, weight, and utilise diagnostic features from different sensory modalities in a goal-directed manner, with these processes changing systematically across development.

A central contribution of this work is the distinction between associative binding and integrative processing in early category learning. Findings from Chapter 3 indicate that both children and adults benefit from consistent audiovisual pairings, with performance declining when learned motion–sound combinations are disrupted. However, the absence of strong developmental differences in our child participants (5-13 years) suggests that these effects are better explained by cross-modal associative learning, the encoding of co-occurring features, rather than fully mature multisensory integration. This interpretation aligns with developmental accounts emphasising sensitivity to statistical regularities across modalities (Lewkowicz & Kraebel, 2004; Sloutsky, 2010) and suggests that early category representations may be multifeatural but not yet optimally integrated. The presence of improved categorisation in our adult population when compared to children may therefore be interpreted as this shift towards mature multisensory representations (Mercier & Cappe, 2020) which could result in enhanced adult performance (Colonius & Diederich, 2020; Ernst & Bulthoff, 2004).

Building on this, Chapter 4 suggests that multisensory object category representations can be unevenly weighted in children, rather than purely modality dominant. Although haptic information exerted a stronger influence on categorisation behaviour, this effect may be understood as reflecting task-dependent cue weighting rather than a fixed haptic dominance in children. Within a reliability-weighting framework (Ernst & Banks, 2002), learners appear to prioritise information that is more diagnostic or salient under the given task conditions. Developmental evidence further suggests that children may

rely more heavily on a single modality when processing demands are high or when integration mechanisms are still maturing (Gori et al., 2008; Nardini et al., 2008). Importantly, the findings do not indicate that visual information is absent from visuohaptic object category representations, but in the current work visual diagnostic cues contributed less strongly to behavioural decisions, perhaps due to the saliency of haptic information. reflecting asymmetric weighting within associative cross-modal or multisensory representations.

Chapter 5 extends these findings by demonstrating that category learning is also shaped by top-down mechanisms governing attention and exploration. The observation that uncertainty enhanced learning performance suggests that learners actively adapt their information-sampling strategies in response to task demands, consistent with active inference accounts (Friston, 2010). This highlights a critical role for uncertainty-driven exploration in category learning, whereby learners seek information that reduces uncertainty and improves predictive accuracy. Furthermore, eye-tracking data revealed a dissociation between overt attention and representational weighting, with diagnostically relevant features not receiving the greatest visual sampling, though behaviourally guiding category decision making. This finding contributes to emerging evidence that category learning is guided by internal models that are only partially reflected in observable overt gaze behaviour (Weichart et al., 2024; Gao et al., 2024; Rehder & Hoffman, 2005).

Taken together, these findings can be situated within a predictive processing framework, in which category learning involves the construction of hierarchical generative models that integrate prior expectations with incoming sensory input (Clark, 2013). From this perspective, development reflects increasing efficiency in how these models are formed and updated. Early in development, category learning appears to be constrained by the precision of sensory input and characterised by broad, similarity-based representations

supported by associative binding (Lewkowicz & Kraebel, 2004). As development progresses, learners become better able to selectively attend to diagnostic features, weight cues according to reliability, and form more abstract, generalisable representations, consistent with transitions from exemplar-based to rule-based processing (Rabi & Minda, 2014; Ramscar & Gitcho, 2007).

Crucially, the present thesis demonstrates that top-down influences on multisensory category learning can interact with bottom-up processes in a non-additive manner. Rather than simply enhancing multisensory integration, top-down mechanisms, including predictive abstraction (Clark, 2013; Friston, 2010), attentional selection (Kruschke, 2001; Nosofsky, 1986), and exploratory control (Friston et al., 2015; Bonawitz et al., 2012), actively shape how sensory information is encoded and used. These processes are supported by the maturation of frontoparietal control systems (Casey et al., 2019), which enable more efficient regulation of attention, hypothesis testing, and uncertainty resolution (Miller & Cohen, 2001).

In sum, this work refines current multisensory categorisation theory by proposing that development involves a shift from bottom-up constrained learning to increasingly top-down guided processing, while remaining fundamentally dependent on the saliency of sensory input. Multisensory experience contributes not only by providing redundant information (Bahrick & Lickliter, 2000; Stein & Meredith, 1993), but by stabilising predictive models and supporting flexible generalisation (Newell et al., 2023; Murray & Shams, 2023). Thus, mature category learning reflects a dynamic balance between accurate perceptual encoding (Viganò et al., 2021) and the efficient deployment of predictive (Clark, 2013), attentional (Kruschke, 2001), and exploratory mechanisms (Friston, 2010), offering a more comprehensive account of how category knowledge is acquired and generalised across development.

Novelty

This thesis extends previous work in both visuohaptic (Broadbent et al., 2020a) and audiovisual (Broadbent et al., 2018b; Roark et al., 2023) category learning by addressing two central questions: (a) how children acquire multisensory categories for complex, three-dimensional (3D) objects, and (b) the nature of the resulting category representations in memory. The overarching goal was to move beyond demonstrations of multisensory facilitation to characterise how sensory information is encoded, associated, and weighted in children's developing category representations. In order to accomplish this, multiple novel stimulus sets were developed using previous work as inspiration (Broadbent et al., 2020) we developed 3 sets of 2 novel alien species and by incorporating 'gamified' experimental paradigms, our experiments aimed to be specifically engaging to our developmental sample, while providing us with rich insights into not only category acquisition but also the object category representations stored in memory.

In Chapter 3, a novel stimulus set was created comprising of correlated dynamic visual and auditory cues. While prior work in children has demonstrated the progressive ability to integrate static audiovisual cues for categorisation (Roark et al., 2023), as well as multisensory facilitation in category learning (Broadbent et al., 2018); the current study extended this to temporally correlated dynamic audiovisual objects. The visual motion of each stimulus was aligned with auditory pitch and temporal duration, allowing examination of whether visual movement–sound temporal correspondence facilitated the formation of integrated category representations. Previous studies of multisensory category learning (Broadbent et al., 2018; Roark et al., 2023) demonstrated audiovisual integration benefits at the level of performance; here, we assessed whether correlated cues are bound in memory as a single representational unit. Both children and adults showed enhanced categorisation accuracy for simultaneously presented auditory, visual, and motion cues, suggesting that

all three dimensions were associated within their object representations. These findings are consistent with evidence that multisensory congruency and temporal alignment enhance learning and retention through enhanced associative binding (Newell et al., 2023; Hillock et al., 2011), and that dynamic correlations play a key role in cue binding during development (Kaganovich, 2012).

Chapters 4 and 5 further extended previous work in visuohaptic category-learning paradigms (Broadbent et al., 2020) by incorporating methodological manipulations drawn from visual category-learning studies of dimensional attention (e.g., Sloutsky & Fisher, 2008; Plebanek & Sloutsky, 2017). We introduced cross-category feature stimuli, objects whose diagnostic features (visual, haptic, or visuohaptic) were switched to that of the opposing category, to investigate which features children relied upon when learning novel categories. This design allowed examination of both perceptual and decisional weighting of cues when multiple diagnostic features were available.

Chapter 2 examined familiar object categorisation using both contextual and object-based audiovisual cues, allowing comparison of the relative benefit of scene context versus object-specific sensory information. Whereas much of the child categorisation literature has focused on unimodal visual tasks (Huang-Pollock et al., 2011; López et al., 1992; Rabi & Minda, 2014; Mareschal & Quinn, 2001), the present work integrated cross-modal and contextual manipulations. A visual scene context was paired with an auditory object cue enabling assessment of whether contextual or object-specific information drives categorisation performance. The finding that congruent contextual cues paired with matching auditory object cues enhanced accuracy indicates that, even in middle childhood, multisensory semantic congruency facilitates object categorisation, a pattern consistent with adult findings on object–scene consistency (Bar, 2004; Davenport & Potter, 2004),

and with the gradual strengthening of top-down predictive mechanisms that integrate context and sensory evidence (Talsma, 2015).

In Chapter 5, we employed an eye-tracking paradigm to explore attentional mechanisms underpinning multisensory category decisions. Although the hypothesised group differences in visual fixation patterns under varying instructional conditions were not fully realised, the paradigm provided a rigorous foundation for assessing the relationship between overt attention and cue weighting. Notably, a dissociation emerged between gaze allocation and the cues that influenced decision-making: children and adults often directed attention to either non-diagnostic features or visual diagnostic features however, both groups weighted visuohaptic cues more heavily in their categorisation judgments. These finding parallels adult work demonstrating that visual attention and decision weighting can diverge (Rehder & Hoffman, 2005) and highlights the need to distinguish attentional allocation from decisional weighting in developing category systems. Such dissociation accords with developmental models of selective attention showing that children's attentional control is broad and diffuse relative to adults' (Plebanek & Sloutsky, 2017; Ng et al., 2024), and that refinement of frontoparietal control networks across childhood supports increasingly selective weighting of relevant sensory information.

Overall, this thesis contributes novel insights into the mechanisms by which children acquire multisensory object categories. Across audiovisual and visuohaptic domains, we demonstrate that multisensory cues are not merely additive sources of information but are integrated into unified representational structures in memory. The developmental trajectories observed across experiments reveal that cue binding, feature weighting, and top-down attentional control interact to shape how multisensory experiences are abstracted into categorical knowledge. By uniting empirical approaches from audiovisual, visuohaptic, and attentional research traditions, this work provides a

comprehensive account of how complex, real-world multisensory object categories emerge across childhood.

Limitations

Although the present thesis offers novel insights into the development of multisensory category learning across childhood, several methodological and conceptual limitations should be acknowledged. These relate primarily to the lack of preregistration, task design and difficulty, sample characteristics, data variability, and measurement criteria for learning thresholds.

Pre-registration and Analytical Transparency

An initial limitation of the present thesis to be addressed is the lack of pre-registering the experiments prior to data collection. Pre-registration is increasingly recognised as best practice in psychological science (Nosek et al., 2015; Munafo et al., 2017), as it enhances transparency, constrains analytic flexibility, and distinguishes confirmatory from exploratory analyses (Open Science Collaboration, 2015; Nosek et al., 2018). Its absence therefore limits the extent to which the present findings can be interpreted as strictly confirmatory. However, each experiment was strictly theory-driven and guided by clearly defined research questions grounded in existing literature which provides some support for their robustness. Nonetheless, analytic decisions (e.g., inclusion criteria or performance thresholds) may have introduced a degree of flexibility in the reported outcomes. This is particularly relevant in developmental research, where variability in performance often necessitates adaptive design and analysis choices.

Future work should incorporate pre-registered designs where feasible, specifying hypotheses, analytic plans, and inclusion criteria in advance. Where flexibility is required, hybrid approaches that combine pre-registered hypotheses with transparent exploratory analyses may be especially appropriate. Addressing this limitation will strengthen the

evidential basis and reproducibility of research on multisensory category learning across development.

Experimental Design and Task Difficulty across participants

Designing an experimental paradigm that is both feasible for a 4-year-old and sufficiently challenging for a 13-year-old is an inherent difficulty in developmental cognitive research (Flavell, 1994; Bjorklund, 2018; Bejjanki et al., 2020). The tasks presented across Chapters 2–5 were intentionally constructed to maintain structural comparability across age groups, yet inevitably the perceptual, cognitive, and attentional demands differed for younger versus older children. In particular, the multisensory category-learning tasks (Chapters 3–4) required children to learn complex 3D stimulus–response mappings involving several correlated sensory dimensions. In Chapter 4, approximately 70 % of our 4–6-year-old participants did not meet the performance criterion for inclusion. While this likely reflects the genuine difficulty of acquiring novel visuohaptic object categories particularly as our stimuli were configured to be complex, it raises questions about representativeness: the subset of children who successfully completed the task demonstrated excellent learning but may represent a particularly high-performing or motivated group. Consequently, the results may overestimate the ease with which children in the general population acquire such multisensory categories. Future work could address this limitation through adaptive designs that titrate task complexity to individual ability levels (e.g., staircase procedures or adaptive block progression), enabling inclusion of a wider range of developmental profiles.

Data Variability in Developmental Research

Children, particularly those in early and middle childhood, are inherently variable participants. Fluctuations in attention, fatigue, motivation, and comprehension can contribute to high intra- and inter-individual variability (Davidson et al., 2006). Despite

extensive piloting, counterbalancing, and exclusion criteria, children's behavioural data often exhibit greater variance than adult data, especially in multisensory paradigms where attention must be divided or flexibly allocated (Talsma, 2015). In the present thesis, the final analyses reflect carefully screened data; however, it is important to recognise that the obtained results represent a subset of children able to sustain task engagement and comply with experimental instructions. Although this is a common constraint in developmental research, it underscores the need for replication using complementary methodologies such as simplified paradigms, passive measures (e.g., looking-time or EEG designs for younger children), or tasks embedded in game-based environments that maintain motivation without compromising rigour (Nacke & Deterding, 2017).

Sample Characteristics and Representativeness

All participating children were recruited from mainstream educational settings and thus represent typically developing samples within standard schooling systems. While this homogeneity supports interpretability, it limits generalisability to children with atypical sensory or learning profiles. The ability to integrate and weight cues has been suggested to differ significantly in children with sensory-processing or neurodevelopmental differences (Molholm et al., 2020; Stevenson et al., 2014). Moreover, the age groupings across chapters were not fully consistent. In Chapters 3 and 4, children were divided into discrete age bands to maximise comparability within feasible sample sizes; in Chapter 2, age was treated as a continuous variable to capture developmental gradients in familiar object categorisation; and in Chapter 5, a narrower age range (8–11 years) was adopted to enable use of eye-tracking. These pragmatic differences were driven by recruitment feasibility and equipment constraints, yet they reduce the degree of direct comparability across experiments.

Originally, the project aimed to test children aged 4–12 years. However, due to changes in national schooling norms and recruitment logistics, the tested range was

adjusted to 5–13 years, with limited inclusion of 4-year-olds. The youngest preschool-age children—who might provide critical insights into the earliest stages of multisensory integration—were therefore underrepresented. Future work could explicitly target this group, for example through partnerships with preschool and Montessori settings, or by adapting paradigms for tablet-based testing and eye-tracking suitable for non-readers. Including these younger ages would enable a more fine-grained characterisation of the emergence of multisensory category formation and could clarify whether developmental transitions observed here reflect gradual or stage-like changes.

Operationalisation of 'Learning' and Criterion Thresholds

Defining a threshold for "category learning" presents a further methodological challenge. In this thesis, different operational definitions were applied across experiments, each selected to best match the category structure and time constraints of developmental testing. In Chapter 3, a sliding-window analysis required at least 75 % accuracy over the previous eight trials; in Chapter 4, a stricter criterion of eight consecutive correct responses was used; and in Chapter 5, a block-level criterion of 75 % accuracy following a minimum number of blocks was adopted. These thresholds were informed by extensive piloting and designed to balance statistical reliability with the practical realities of maintaining children's attention over extended sessions. Nonetheless, varying the definition of "learning" may have influenced observed success rates and comparability across studies. The stricter criterion in Chapter 4, for instance, may partially account for the high exclusion rate. Although the final, block-level approach (Chapter 5) provides a robust compromise between sensitivity and feasibility, future research might benefit from hierarchical Bayesian models or drift-diffusion frameworks that estimate learning rate and decision precision continuously, rather than imposing discrete pass/fail criteria (Lee & Wagenmakers, 2014; Zhang & Rowe, 2014).

Taken together, these limitations highlight both the complexity and the necessity of studying multisensory category learning in development. The challenges encountered—task calibration across ages, participant variability, and defining robust learning metrics—are integral to the field rather than specific shortcomings of this project. Nevertheless, they provide concrete directions for future research. Adaptive task design expanded sampling to include preschool populations and diverse sensory profiles, and the adoption of model-based learning metrics will further advance understanding of how children acquire and represent multisensory object categories across development.

Conclusion

The research presented in this thesis underlines how fundamental multisensory experience is in daily life and thus to the nature of object perception. It further examined how the ability to combine and utilise information across sensory modalities to aid perceptual judgements evolves throughout childhood. Across the examined age range of 4-13 years, children demonstrated increasing sophistication in how they utilised visual, auditory, and haptic information, and their combinations, to acquire and employ object category knowledge. These findings collectively reveal that multisensory object categories become more robust through development, reflecting with our behavioural findings the maturational changes and experiential calibration that underscores this developmental period.

A central contribution of this work lies in demonstrating that children are capable of utilising multisensory information and identifying multisensory redundancies in cross-modal categorisation tasks. Moreover, the empirical findings emphasise the task-specific emergence of multisensory facilitation, with limited evidence for benefits for category learning. Nevertheless, where benefits emerge with age and multisensory exposure is in children's ability to generalise their category learning. This supports the robustness and

generalisability of category representations after multisensory exposure. The observed cross-modal transfer effects indicate that learning in one sensory domain can influence performance in another, providing evidence for shared representational structures that underpin flexible object perception. This capacity for cross-modal generalisation reflects an increasingly coordinated sensory system that supports more efficient and versatile category learning with increasing age.

Taken together, these results contribute to a more nuanced understanding of the developmental trajectory of multisensory cognitive development. In doing so, it underscores the importance of considering the dynamic and task specific nature of multisensory benefits to the formation and refinement of category knowledge in children.

References

Abdala, C. (2000). Distortion product otoacoustic emission (2f1-f2) amplitude growth in human adults and neonates. *The Journal of the Acoustical Society of America*, *107*(1), 446-456. https://doi.org/10.1121/1.428315

Abdala, C., & Folsom, R. C. (1995). The development of frequency resolution in humans as revealed by the auditory brain-stem response recorded with notched-noise masking. *The Journal of the Acoustical Society of America*, *98*(2), 921-930. https://doi.org/10.1121/1.414350

Adams, W. J., et al. (2016). The development of audiovisual integration for temporal perception. *PLoS Computational Biology, 12*(4), e1004865. https://doi.org/10.1371/journal.pcbi.1004865

Aguiar, A., & Baillargeon, R. (1999). 2.5-month-old infants' reasoning about when objects should and should not be occluded. *Cognitive Psychology*, *39*(2), 116-157. https://doi.org/10.1006/cogp.1999.0717

AlAhmed, F., Rau, A., & Wallraven, C. (2023). Visuo-haptic processing of unfamiliar shapes: Comparing children and adults. *Plos one*, *18*(10), e0286905. https://doi.org/10.1371/journal.pone.0286905

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*(3), 257–262. https://doi.org/10.1016/j.cub.2004.01.029

Alexander, J. M., Johnson, K. E., & Schreiber, J. B. (2002). Knowledge is not everything: Analysis of children's performance on a haptic comparison task. *Journal of Experimental Child Psychology*, *82*(4), 341-366. https://doi.org/10.1016/S0022-0965(02)00100-5

References

Allen, P., & Wightman, F. (1994). Psychometric functions for children's detection of tones in noise. *Journal of Speech, Language, and Hearing Research*, *37*(1), 205-215. https://doi.org/10.1044/jshr.3701.205

Almasi, R. C., & Behrmann, M. (2021). Subcortical regions of the human visual system do not process faces holistically. *Brain and Cognition*, *151*, 105726. https://doi.org/10.1016/j.bandc.2021.105726

Althaus, N., & Mareschal, D. (2014). Labels direct infants' attention to commonalities during novel category learning. *PloS one*, *9*(7), e99670. https://doi.org/10.1371/journal.pone.0099670

Alvarado, J. C., Rowland, B. A., Stanford, T. R., & Stein, B. E. (2008). A neural network model of multisensory integration also accounts for unisensory integration in superior colliculus. *Brain Research*, *1242*, 13-23.https://doi.org/10.1016/j.brainres.2008.03.074

Amedi, A., Malach, R., Hendler, T., Peled, S., & Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. *Nature Neuroscience, 4*(3), 324–330. https://doi.org/10.1038/85201

Amedi, A., Stern, W. M., Camprodon, J. A., et al. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nature Neuroscience, 10*(6), 687–689. https://doi.org/10.1038/nn1912

Amso, D., Haas, S., & Markant, J. (2014). An eye tracking investigation of developmental change in bottom-up attention orienting to faces in cluttered natural scenes. *PloS one*, *9*(1), e85701. https://doi.org/10.1371/journal.pone.0085701

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98*(3), 409–429. https://doi.org/10.1037/0033-295X.98.3.409

References

Arnott, S. R., Binns, M. A., Grady, C. L., & Alain, C. (2004). Assessing the auditory dual-pathway model in humans. *NeuroImage, 22*(1), 401–408. https://doi.org/10.1016/j.neuroimage.2004.01.014

Arterberry, M. E., & Bornstein, M. H. (2002). Infant perceptual and conceptual categorization: The roles of static and dynamic stimulus attributes. *Cognition*, *86*(1), 1-24. https://doi.org/10.1016/S0010-0277(02)00108-7

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology, 56*, 149–178. https://doi.org/10.1146/annurev.psych.56.091103.070217

Aslin, R. N., & Banks, M. S. (1978). Early visual experience in humans: Evidence for a critical period in the development of binocular vision. In *Psychology: From research to practice* (pp. 227-239). Boston, MA: Springer US.

Aslin, R. N., & Smith, L. B. (1988). Perceptual development. *Annual Review of Psychology*, *39*, 435–473. https://doi.org/10.1146/annurev.ps.39.020188.002251

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Perey, A. J. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. *Child Development*, *52*(4), 1135. https://doi.org/10.2307/1129499

Atkinson, J. (2017). The Davida Teller Award Lecture, 2016: visual brain development: a review of "dorsal stream vulnerability"—motion, mathematics, amblyopia, actions, and attention. *Journal of Vision*, *17*(3), 26-26. https://doi.org/10.1167/17.3.26

Atkinson, J., Anker, S., Rae, S., Weeks, F., Braddick, O., & Rennie, J. (2002). Cortical visual evoked potentials in very low birthweight premature infants. *Archives of*

Disease in Childhood-Fetal and Neonatal Edition, *86*(1), F28-F31. https://doi.org/10.1136/fn.86.1.F28

Atkinson, J., Braddick, O., & Braddick, F. (1974). Acuity and contrast sensivity of infant vision. *Nature*, *247*(5440), 403-404. https://doi.org/10.1038/247403a0

Augustine, E., Smith, L. B., & Jones, S. S. (2011). Parts and relations in young children's shape-based object recognition. *Journal of Cognition and Development*, *12*(4), 556-572. https://doi.org/10.1080/15248372.2011.560586

Ayzenberg, V., & Behrmann, M. (2022). Does the brain's ventral visual pathway compute object shape?. *Trends in Cognitive Sciences*, *26*(12), 1119-1132. https://doi.org/10.1016/j.tics.2022.09.019

Ayzenberg, V., & Behrmann, M. (2024). Development of visual object recognition. *Nature Reviews Psychology, 3*(2), 73-90. https://doi.org/10.1038/s44159-023-00266-w

Ayzenberg, V., & Lourenco, S. (2022). Perception of an object's global shape is best described by a model of skeletal structure in human infants. *elife*, *11*, e74943. https://doi.org/10.7554/eLife.74943

Bahrick, L. E. (2001). Increasing specificity in perceptual development: Infants' detection of nested levels of multimodal stimulation. *Journal of Experimental Child Psychology, 79*(3), 253–270. https://doi.org/10.1006/jecp.2000.2588

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, *36*(2), 190-201. https://doi.org/10.1037/0012-1649.36.2.190

Bahrick, L. E., & Lickliter, R. (2004). Infants' perception of rhythm and tempo in unimodal and multimodal stimulation: A developmental test of the intersensory

redundancy hypothesis. *Cognitive, Affective, & Behavioral Neuroscience, 4*(2), 137–147. https://doi.org/10.3758/CABN.4.2.137

Bahrick, L. E., Hernandez-Reif, M., & Flom, R. (2005). The development of infant learning about specific face-voice relations. *Developmental Psychology*, *41*(3), 541. https://doi.org/10.1037/0012-1649.41.3.541

Bahrick, L. E., Lickliter, R., & Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science, 13*(3), 99–102. https://doi.org/10.1111/j.0963-7214.2004.00283.x

Bahrick, L. E., Lickliter, R., & Flom, R. (2006). Up versus down: The role of intersensory redundancy in infants' sensitivity to tempo and rhythm. *Developmental Psychology, 42*(3), 568–578. https://doi.org//10.1207/s15327078in0901_4

Baillargeon, R., Wu, D., Yuan, S., Li, J., & Luo, Y. (2009). Young infants' expectations about self-propelled objects. *The Origins of Object Knowledge*, 285-352. https://doi.org/10.1093/acprof:oso/9780199216895.003.0012

Baker, J. M., & Jordan, K. E. (2015). The influence of multisensory cues on representation of quantity in children. In *Mathematical cognition and learning* (Vol. 1, pp 277-301). Elsevier. https://doi.org/10.1016/B978-0-12-420133-0.00011-9

Baker, T. J., Tse, J., Gerhardstein, P., & Adler, S. A. (2008). Contour integration by 6-month-old infants: discrimination of distinct contour shapes. *Vision Research*, *48*(1), 136-148. https://doi.org/10.1016/j.visres.2007.10.021

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617-629. https://doi.org/10.1038/nrn1476

References

Bardi L, Regolin L, Simion F. (2011). Biological motion preference in humans at birth: role of dynamic and configural properties. *Developmental Science*. *14*(2) 353-9. https://doi.org/10.1111/j.1467-7687.2010.00985.x.

Bargones, J. Y., Werner, L. A., & Marean, G. C. (1995). Infant psychometric functions for detection: Mechanisms of immature sensitivity. *The Journal of the Acoustical Society of America*, *98*(1), 99-111. https://doi.org/10.1121/1.414446

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barsalou, L. W., Niedenthal, P. M., Barbey, A. K., & Ruppert, J. A. (2003). Social embodiment. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory, 43, pp. 43–92. Elsevier Science.*

Barutchu, A., Crewther, D., & Crewther, S. (2009). The race that precedes coactivation: Development of multisensory facilitation in children. *Developmental Science, 12*(3), 464–473. **https://doi.org/10.1111/j.1467-7687.2008.00782.x**

Barutchu, A., Danaher, J., Crewther, S. G., Innes-Brown, H., Shivdasani, M. N., & Paolini, A. G. (2010). Audiovisual integration in noise by children and adults. *Journal of Experimental Child Psychology*, *105*(1-2), 38-50. https://doi.org/10.1016/j.jecp.2009.08.005

Bastianello, T., Keren-Portnoy, T., Majorano, M., & Vihman, M. (2022). Infant looking preferences towards dynamic faces: A systematic review. *Infant Behavior and Development*, *67*, 101709. https://doi.org/10.1016/j.infbeh.2022.101709

Batty, M., & Taylor, M. J. (2002). Visual categorization during childhood: An ERP study. *Psychophysiology, 39*, 482–490. https://doi.org/10.1017/S0048577202010764

References

Bauer, P. J., & Mandler, J. M. (1989). Taxonomies and triads: Conceptual organization in one-to two-year-olds. *Cognitive Psychology*, *21*(2), 156-184. https://doi.org/10.1016/0010-0285(89)90006-6

Bavelier, D., Dye, M. W., & Hauser, P. C. (2006). Do deaf individuals see better? *Trends in Cognitive Sciences, 10*(11), 512–518. https://doi.org/10.1016/j.tics.2006.09.006

Behl-Chadha, G. (1996). Basic-level and superordinate-like categorical representations in early infancy. *Cognition*, *60*(2), 105-141. https://doi.org/10.1016/0010-0277(96)00706-8

Bejjanki, V. R., Randrup, E. R., & Aslin, R. N. (2020). Young children combine sensory cues with learned information in a statistically efficient manner: But task complexity matters. *Developmental Science*, *23*(3), e12912. https://doi.org/10.1111/desc.12912

Belin, P., & Zatorre, R. J. (2000). 'What', 'where' and 'how' in auditory cortex. *Nature Neuroscience, 3*(10), 965–966. https://doi.org/10.1038/79890

Bendixen, A., Háden, G. P., Németh, R., Farkas, D., Török, M., & Winkler, I. (2015). Newborn infants detect cues of concurrent sound segregation. *Developmental Neuroscience*, *37*(2), 172-181. https://doi.org/10.1159/000370237

Bennett, M. R., & Hacker, P. M. S. (2006). *Philosophical foundations of neuroscience*. Blackwell.

Berger, C., & Donnadieu, S. (2008). Visual/auditory processing and categorization preferences in 5-year-old children and adults. *Current Psychology Letters, 24*(1), 41–57. https://doi.org/10.4000/cpl.3673

Berkeley, G. 1983. Philosophical Works including the Works on Vision. London: Dent.

References

Berland, A., Bruckert, L., Benaroyo, L., & Gaillard, V. (2015). Perception of everyday sounds: A developmental study of auditory categorization. *Frontiers in Psychology, 6*, 1317. https://doi.org/10.1371/journal.pone.0115557

Berland, A., Gaillard, P., Guidetti, M., & Barone, P. (2015). Perception of everyday sounds: A developmental study of a free sorting task. *PLoS One*, *10*(2), e0115557. https://doi.org/10.1371/journal.pone.0115557

Bertenthal, B. I. (1993). Infants' perception of biomechanical motions: Intrinsic image and knowledge-based constraints. In C. Granrud (Ed.), *Visual perception and cognition in infancy* (pp. 175–214). Lawrence Erlbaum Associates, Inc.

Bertenthal, B. I., Proffitt, D. R., & Cutting, J. E. (1984). Infant sensitivity to figural coherence in biomechanical motions. *Journal of Experimental Child Psychology*, *37*(2), 213-230. https://doi.org/10.1016/0022-0965(84)90001-8

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P. W., Kennedy, L. J., & Mehler, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General*, *117*(1), 21. https://doi.org/10.1037/0096-3445.117.1.21

Bhatt, R. S., & Waters, S. E. (1998). Perception of three-dimensional cues in infancy: A developmental analysis. *Child Development, 69*(5), 1249–1260. https://doi.org/10.1006/jecp.1998.2458

Bidet-Ildei, C., Kitromilides, E., Orliaguet, J. P., Pavlova, M., & Gentaz, E. (2014). Preference for point-light human biological motion in newborns: contribution of translational displacement. *Developmental Psychology*, *50*(1), 113. https://doi.org/10.1037/a0032956

References

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*(2), 115–147. https://doi.org/10.1037/0033-295X.94.2.115

Biederman, I., & Bar, M. (1999). One-shot viewpoint invariance in matching novel objects. *Vision Research, 39*(17), 2885–2899. https://doi.org/10.1016/S0042-6989(98)00309-5

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*(2), 143-177. https://doi.org/10.1016/0010-0285(82)90007-X

Bigelow, A. E. (1981). Children's tactile identification of miniaturized common objects. *Developmental Psychology, 17*(1), 111–114. https://doi.org/10.1037/0012-1649.17.1.111

Bingham, G. P., & Muchisky, M. M. (2018). “Center of mass perception”: affordances as dispositions determined by dynamics. In *Global perspectives on the ecology of human-machine systems* (pp. 359-395). CRC Press.

Birnholz, J. C., & Benacerraf, B. R. (1983). The development of human fetal hearing. *Science*, *222*(4623), 516-518. https://doi.org/10.1126/science.6623091

Bizley, J. K., & Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nature Reviews Neuroscience, 14*(10), 693–707. https://doi.org/10.1038/nrn3565

Bjorklund, D. F. (1985). The role of conceptual knowledge in the development of organization in children’s memory. In *Basic Processes in Memory Development: Progress in Cognitive Development Research* (pp. 103-142). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4613-9541-6_3

References

Bjorklund, D. F. (1987). How age changes in knowledge base contribute to the development of children's memory: An interpretive review. *Developmental Review*, *7*(2), 93-130. https://doi.org/10.1016/0273-2297(87)90007-4

Blasi, A., Mercure, E., Lloyd-Fox, S., Thomson, A., Brammer, M., Sauter, D., Deeley, Q., Barker, G.J., Renvall, V., Deoni, S., & Murphy, D. G. (2011). Early specialization for voice and emotion processing in the infant brain. *Current Biology*, *21*(14), 1220-1224. https://doi.org/10.1016/j.cub.2011.06.009

Blewitt, P. (1994). Understanding categorical hierarchies: The earliest levels of skill. *Child Development*, *65*(5), 1279-1298. https://doi.org/10.1111/j.1467-8624.1994.tb00817.x

Bleyenheuft, Y., Cols, C., Arnould, C., & Thonnard, J. L. (2006). Age-related changes in tactile spatial resolution from 6 to 16 years old. *Somatosensory & Motor Research*, *23*(3–4), 83–87. https://doi.org/10.1080/08990220600816440

Bogfjellmo, L. G., Bex, P. J., & Falkenberg, H. K. (2014). The development of global motion discrimination in school aged children. *Journal of Vision*, *14*(2), 19-19. https://doi.org/10.1167/14.2.19

Bomba, P. C. (1984). The development of orientation categories between 2 and 4 months of age. *Journal of Experimental Child Psychology*, *37*(3), 609-636. https://doi.org/10.1016/0022-0965(84)90080-8

Bomba, P. C., & Siqueland, E. R. (1983). The nature and structure of infant form categories. *Journal of Experimental Child Psychology*, *35*(2), 294-328. https://doi.org/10.1016/0022-0965(83)90085-1

Bonawitz, E. B., van Schijndel, T. J., Friel, D., & Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology*, *64*(4), 215-234. https://doi.org/10.1016/j.cogpsych.2011.12.002

References

Bonin, P., Gelin, M., Laroche, B., Méot, A., & Bugaiska, A. (2015). The “how” of animacy effects in episodic memory. *Experimental Psychology*, *62*(6), https://doi.org/10.1027/1618-3169/a000308

Bornstein, M. H., & Arterberry, M. E. (2010). The development of object categorization in young children: hierarchical inclusiveness, age, perceptual attribute, and group versus individual analyses. *Developmental Psychology*, *46*(2), 350 –365. https://doi.org/10.1037/a0018411

Bornstein, M. H., Kessen, W., & Weiskopf, S. (1976). Color vision and hue categorization in young human infants. *Journal of Experimental Psychology: Human Perception and Performance, 2*(1), 115–129. https://doi.org/10.1037/0096-1523.2.1.115

Bova, S. M., Fazzi, E., Giovenzana, A., Montomoli, C., Signorini, S. G., Zoppello, M., & Lanzi, G. (2007). The development of visual object recognition in school-age children. *Developmental Neuropsychology*, *31*(1), 79-102. https://doi.org/10.1080/87565640709336888

Bowman, C. R., Iwashita, T., & Zeithamova, D. (2020). Tracking prototype and exemplar representations in the brain. *eLife, 9*, e59360. https://doi.org/10.7554/eLife.59360

Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 556. https://doi.org/10.1037//0096-1523.15.3.556

Braddick, O. (1996). Binocularity in infancy. *Eye*, *10*(2), 182-188. https://doi.org/10.1038/eye.1996.45

Braddick, O. J., & Atkinson, J. (2009). Infants’ sensitivity to motion and temporal change. *Optometry and Vision Science*, *86*(6), 577-582. https://doi.org/10.1097/OPX.0b013e3181a76e84

References

Braddick, O. J., & Atkinson, J. (2011). Development of human visual function. *Vision Research*, *51*(13), 1588-1609. https://doi.org/10.1016/j.visres.2011.02.018

Braddick, O. J., Wattam-Bell, J., & Atkinson, J. (1986). Orientation-specific cortical responses develop in early infancy. *Nature*, *320*(6063), 617-619. https://doi.org/10.1038/320617a0

Brandwein, A. B., Foxe, J. J., Russo, N., Altschuler, T., Gomes, H., & Molholm, S. (2011). The development of audiovisual multisensory integration across childhood and early adolescence: A high-density electrical mapping study. *Cerebral Cortex, 21*(5), 1042–1055. https://doi.org/10.1093/cercor/bhq170

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press.

Bremner, A. J., Lewkowicz, D. J., & Spence, C. (2008). The multisensory approach to development. In A. J. Bremner, D. J. Lewkowicz, & C. Spence (Eds.), *Multisensory development* (pp. 1–26). Oxford University Press.

Broadbent, H. J., Osborne, T., Mareschal, D., & Kirkham, N. Z. (2019). Withstanding the test of time: Multisensory cues improve the delayed retention of incidental learning. *Developmental Science*, *22*(1), e12726. https://doi.org/10.1111/desc.12726

Broadbent, H. J., Osborne, T., Rea, M., Peng, A., Mareschal, D., & Kirkham, N. Z. (2018a). Incidental category learning and cognitive load in a multisensory environment across childhood. *Developmental Psychology*, *54*(6), 1020. https://doi.org/10.1037/dev0000472

Broadbent, H. J., White, H., Mareschal, D., & Kirkham, N. Z. (2018b). Incidental learning in a multisensory environment across childhood. *Developmental Science*, *21*(2), e12554. https://doi.org/10.1111/desc.12554

References

Broadbent, H., Osborne, T., Kirkham, N., & Mareschal, D. (2020a). Touch and look: The role of visual-haptic cues for categorical learning in primary school children. *Infant and Child Development*, *29*(2), e2168. https://doi.org/10.1002/icd.2168

Broadbent, H., Osborne, T., Mareschal, D., & Kirkham, N. (2020b). Are two cues always better than one? The role of multiple intra-sensory cues compared to multi-cross-sensory cues in children's incidental category learning. *Cognition*, *199*, 104202. https://doi.org/10.1016/j.cognition.2020.104202

Brown, A. M., & Yamamoto, M. (1986). Visual acuity in newborn and preterm infants measured with grating acuity cards. *American Journal of Ophthalmology*, *102*(2), 245-253. https://doi.org/10.1016/0002-9394(86)90153-4

Brunel, L., Goldstone, R. L., Vallet, G., Riou, B., & Versace, R. (2013). When seeing a dog activates the bark. *Experimental Psychology*. https://doi.org/10.1027/1618-3169/a000176

Bryant, P. E., & Raz, I. (1975). Visual and tactual perception of shape by young children. *Developmental Psychology*, *11*(4), 525. https://doi.org/10.1037/h0076663

Burkhalter, A. (1993). Development of forward and feedback connections between areas V1 and V2 of human visual cortex. *Cerebral Cortex, 3*(5), 476–487. https://doi.org/10.1093/cercor/3.5.476

Burnham, D., & Dodd, B. (2004). Auditory–visual speech perception development in infants: The role of experience. In D. G. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 313–329). MIT Press. https://doi.org/10.1002/dev.20032

References

Burr, D., & Gori, M. (2012). Multisensory integration develops late in humans. In M. M. Murray & M. T. Wallace (Eds.), *The neural bases of multisensory processes* (pp. 345–362). CRC Press/Taylor & Francis.

Bushnell, E. W., & Boudreau, J. P. (1993). Motor development and the mind: The potential role of motor abilities as a determinant of aspects of perceptual development. *Child Development*, *64*(4), 1005-1021. https://doi.org/10.1111/j.1467-8624.1993.tb04184.x

Bushnell, E. W., & Boudreau, J. P. (1998). Exploring and exploiting objects with the hands during infancy.

Bushnell, E. W., & Baxt, C. (1999). Children's haptic and cross-modal recognition with familiar and unfamiliar objects. *Journal of Experimental Psychology: Human Perception and Performance*, 25(6), 1867–1881. https://doi.org/10.1037//0096-1523.25.6.1867

Calce, R. P., Rekow, D., Barbero, F. M., Kiseleva, A., Talwar, S., Leleu, A., & Collignon, O. (2024). Voice categorization in the four-month-old human brain. *Current Biology*, *34*(1), 46-55. https://doi.org/10.1016/j.cub.2023.11.042

Calcus, A. (2024). Development of auditory scene analysis: a mini-review. *Frontiers in Human Neuroscience*, *18*, 1352247. https://doi.org/10.3389/fnhum.2024.1352247

Calvert, G. A., Spence, C., & Stein, B. E. (2004). *The handbook of multisensory processes*. MIT Press. https://doi.org/10.7551/mitpress/3422.001.0001

Candy, T. R., & Banks, M. S. (1999). Use of an early nonlinearity to measure optical and receptor resolution in the human neonate. *Vision Research*, *39*(20), 3386-3398. https://doi.org/10.1016/S0042-6989(99)00035-8

References

Canfield, R. L., & Haith, M. M. (1991). Young infants' visual expectations for symmetric and asymmetric stimulus sequences. *Developmental Psychology*, *27*(2), 198. https://doi.org/10.1037/0012-1649.27.2.198

Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain: The animate-inanimate distinction. *Journal of Cognitive Neuroscience*, *10*(1), 1-34. https://doi.org/10.1162/089892998563752

Casey, B. J., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., & Dale, A. M. (2019). The adolescent brain cognitive development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience, 32*, 43–54. https://doi.org/10.1016/j.dcn.2018.03.001

Casey, B. J., Tottenham, N., Liston, C., & Durston, S. (2005). Imaging the developing brain: What have we learned about cognitive development? *Trends in Cognitive Sciences, 9*(3), 104–110. https://doi.org/10.1016/j.tics.2005.01.011

Casile, A., Caggiano, V., & Ferrari, P. F. (2011). The mirror neuron system: A fresh view. *Neuroscientist, 17*(5), 524–538. https://doi.org/10.1177/1073858410392239

Cassia, V. M., Simion, F., Milani, I., & Umiltà, C. (2002). Dominance of global visual properties at birth. *Journal of Experimental Psychology: General, 131*(3), 398–411. https://doi.org/10.1037/0096-3445.131.3.398

Catherwood, D. (1993). The robustness of infant haptic memory: Testing its capacity to withstand delay and haptic interference. *Child Development*, *64*(3), 702-710.

Chen, W., & Cheung, O. S. (2019). Distinguishing the effects of object-scene association strength and real-world object size in scene priming. *Journal of Vision*, *19*(10), 58-58. https://doi.org/10.1167/19.10.58

Chen, Y. C., & Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked

pictures. *Cognition*, *114*(3), 389-404. https://doi.org/10.1016/j.cognition.2009.10.012

Chen, Y. C., & Spence, C. (2011). Crossmodal semantic priming by naturalistic sounds and spoken words enhances visual sensitivity. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(5), 1554. https://doi.org/10.1037/a0024329

Chen, Y. C., & Spence, C. (2017). Assessing the role of the 'unity assumption' multisensory integration: A review. *Frontiers in Psychology*, *8*, 445. https://doi.org/10.3389/fpsyg.2017.00445

Chen, Y. C., Lewis, T. L., Shore, D. I., Spence, C., & Maurer, D. (2018). Developmental changes in the perception of visuotactile simultaneity. *Journal of Experimental Child Psychology*, *173*, 304-317. https://doi.org/10.1016/j.jecp.2018.04.014

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, *1*(5), 351-353. https://doi.org/10.1038/1561

Cheour-Luhtanen, M., Alho, K., Kujala, T., Sainio, K., Reinikainen, K., Renlund, M., Aaltonen, O., Eerola, O., & Näätänen, R. (1995). Mismatch negativity indicates vowel discrimination in newborns. *Hearing Research*, *82*(1), 53-58. https://doi.org/10.1016/0378-5955(94)00164-L

Chow, J. K., Palmeri, T. J., & Gauthier, I. (2022). Haptic object recognition based on shape relates to visual object recognition ability. *Psychological Research*, *86*(4), 1262-1273. https://doi.org/10.1007/s00426-021-01560-z

Chung, Y. H., Lee, J., & Chong, S. C. (2023). The role of motion in visual working memory for dynamic stimuli: More lagged but more precise representations of

moving objects. *Attention, Perception, & Psychophysics, 85*(2), 438-451. https://doi.org/10.3758/s13414-022-02635-8

Cirelli, L. K., Talukder, L. S., & Kragness, H. E. (2024). Infant attention to rhythmic audiovisual synchrony is modulated by stimulus properties. *Frontiers in Psychology*, *15*, 1393295. https://doi.org/10.3389/fpsyg.2024.1393295

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181–204. https://doi.org/0.1017/S0140525X12000477

Clifton, R. K. (1992). The Development of Spatial Hearing in Human Infants. In L. A. Werner & E. W. Rubel (Eds.), *Developmental psychoacoustics* (pp. 135–157). American Psychological Association. https://doi.org/10.1037/10119-005

Clifton, R. K., Morrongiello, B. A., Kulig, J. W., & Dowd, J. M. (1981). Newborns' orientation toward sound: Possible implications for cortical development. *Child Development*, 833-838. https://doi.org/10.2307/1129084

Coats, R. O. A., Britten, L., Utley, A., & Astill, S. L. (2015). Multisensory integration in children with Developmental Coordination Disorder. *Human Movement Science*, *43*, 15-22. https://doi.org/10.1016/j.humov.2015.06.011

Cohen, L. B., & Younger, B. A. (1984). Infant perception of angular relations. *Child Development, 55*(2), 459–471. https://doi.org/10.1016/S0163-6383(84)80021-1

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics*, *16*(2), 409-412. https://doi.org/10.3758/BF03203962

Collins, A. A., & Gescheider, G. A. (1989). The measurement of loudness in individual children and adults by absolute magnitude estimation and cross-modality matching. *The Journal of the Acoustical Society of America*, *85*(5), 2012-2021. https://doi.org/10.1121/1.397854

References

Colonius, H., & Diederich, A. (2020). Formal models and quantitative measures of multisensory integration: a selective overview. *European Journal of Neuroscience, 51*(5), 1161-1178. https://doi.org/10.1111/ejn.13813

Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review, 112*(2), 347–382. https://doi.org/10.1037/0033-295X.112.2.347

Coogan, T. A., & Van Essen, D. C. (1996). Development of connections within and between areas V1 and V2 of macaque monkey visual cortex. *Journal of Comparative Neurology, 372*(3), 327–342. https://doi.org/10.1002/(SICI)1096-9861(19960826)372:3<327::AID-CNE1>3.0.CO;2-4

Craddock, M., & Lawson, R. (2009). Size-sensitive perceptual representations underlie visual and haptic object recognition. *PLoS ONE, 4*(11), e8009. https://doi.org/10.1371/journal.pone.0008009

Cuppini, C., Ursino, M., Magosso, E., Rowland, B. A., & Stein, B. E. (2011). A computational model of the SC multisensory neurons: Integrative capabilities, maturation, and plasticity. *i-Perception*, *2*(8), 815-815. https://doi.org/10.1068/ic815

Da Costa, S., Bourquin, N. M. P., Knebel, J. F., Saenz, M., Van der Zwaag, W., & Clarke, S. (2015). Representation of sound objects within early-stage auditory areas: a repetition effect study using 7T fMRI. *PloS one*, *10*(5), e0124072. https://doi.org/10.1371/journal.pone.0124072

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

References

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science, 15*(8), 559–564. https://doi.org/10.1111/j.0956-7976.2004.00719.x

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), 2037-2078. https://doi.org/10.1016/j.neuropsychologia.2006.02.006

Deák, G. O., & Bauer, P. J. (1996). The dynamics of preschoolers' categorization choices. *Child Development*, *67*(3), 740-767. https://doi.org/10.1111/j.1467-8624.1996.tb01762.x

Deen, B., Richardson, H., Dilks, D. D., Takahashi, A., Keil, B., Wald, L. L., Kanwisher, N., & Saxe, R. (2017). Organization of high-level visual cortex in human infants. *Nature Communications*, *8*(1), 13995. https://doi.org/10.1038/ncomms13995

Dekker, T., Mareschal, D., Sereno, M. I., & Johnson, M. H. (2011). Dorsal and ventral stream activation and object recognition performance in school-age children. *Neuroimage*, 57(3), 659–670. https://doi.org/10.1016/j.neuroimage.2010.11.005

Deng, W. S., & Sloutsky, V. M. (2015). The development of categorization: effects of classification and inference training on category representation. *Developmental Psychology*, *51*(3), 392. https://doi.org/10.1037/a0038749

Deng, W. S., & Sloutsky, V. M. (2016). Selective attention, diffused attention, and the development of categorization. *Cognitive Psychology*, *91*, 24-62. https://doi.org/10.1016/j.cogpsych.2016.09.002

References

Denner, B., & Cashdan, S. (1967). Sensory processing and the recognition of forms in nursery-school children. *British Journal of Psychology*, *58*(1-2), 101-104. https://doi.org/10.1111/j.2044-8295.1967.tb01062.x

Dodd, B. (1979). Lip reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology, 11*(4), 478–484. https://doi.org/10.1016/0010-0285(79)90021-5

Downing, H. C., Barutchu, A., & Crewther, S. G. (2015). Developmental trends in the facilitation of multisensory objects with distractors. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2014.01559

Downing, H. C., Barutchu, A., & Crewther, S. G. (2015). Developmental trends in the facilitation of multisensory objects with distractors. *Frontiers in Psychology*, *5*, 1559. https://doi.org/10.3389/fpsyg.2014.01559

Duh, S., & Wang, S. H. (2014). Infants detect changes in everyday scenes: The role of scene gist. *Cognitive Psychology*, *72*, 142-161. https://doi.org/10.1016/j.cogpsych.2014.03.001

Ecker, U. K. H., Maybery, M., & Zimmer, H. D. (2013). Binding of intrinsic and extrinsic features in working memory. *Journal of Experimental Psychology: General, 142*(1), 218–234. https://doi.org/10.1037/a0028732

Eggermont, J. J., Brown, D. K., Ponton, C. W., & Kimberley, B. P. (1996). Comparison of distortion product otoacoustic emission (DPOAE) and auditory brain stem response (ABR) traveling wave delay measurements suggests frequency-specific synapse maturation. *Ear and hearing*, *17*(5), 386-394. https://doi.org/10.1097/00003446-199610000-00004

Ehret, G. (1997). The auditory cortex. *Journal of Comparative Physiology A*, *181*(6), 547-557. https://doi.org/10.1007/s003590050139

References

Eilers, R. E., & Minifie, F. D. (1975). Fricative discrimination in early infancy. *Journal of speech and Hearing Research*, *18*(1), 158-167. https://doi.org/10.1044/jshr.1801.158

Eimas, P. D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the [rl] distinction by young infants. *Perception & Psychophysics*, *18*(5), 341-347. https://doi.org/10.3758/BF03211210

Eimas, P. D., & Quinn, P. C. (1994). Studies on the formation of perceptually based basic-level categories in young infants. *Child Development*, *65*(3), 903-917. https://doi.org/10.1111/j.1467-8624.1994.tb00792.x

Eimas, P. D., Quinn, P. C., & Cowan, P. (1994). Development of exclusivity in perceptually based categories of young infants. *Journal of Experimental Child Psychology*, *58*(3), 418-431. https://doi.org/10.1006/jecp.1994.1043

Elfenbein, J. L., Small, A. M., & Davis, J. M. (1993). Developmental patterns of duration discrimination. *Journal of Speech, Language, and Hearing Research*, *36*(4), 842-849. https://doi.org/10.1044/jshr.3604.842

Ellemberg, D., Lewis, T. L., Maurer, D., Brar, S., & Brent, H. P. (2002). Better perception of global motion after monocular than after binocular deprivation. *Vision Research, 42*(2), 169–179. https://doi.org/10.1016/S0042-6989(01)00278-4

Ellis, C. T., Yates, T. S., Skalaban, L. J., Bejjanki, V. R., Arcaro, M. J., & Turk-Browne, N. B. (2021). Retinotopic organization of visual cortex in human infants. *Neuron*, *109*(16), 2616-2626. https://doi.org/10.1016/j.neuron.2021.06.004

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48*(1), 71–99. https://doi.org/10.1016/0010-0277(93)90058-4

References

Engel, L. R., Frum, C., Puce, A., Walker, N. A., & Lewis, J. W. (2009). Different categories of living and non-living sound-sources activate distinct cortical networks. *Neuroimage*, *47*(4), 1778-1791. https://doi.org/10.1016/j.neuroimage.2009.05.041

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*(6870), 429–433. https://doi.org/10.1038/415429a

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8*(4), 162–169. https://doi.org/10.1016/j.tics.2004.02.002

Espinosa, J. S., & Stryker, M. P. (2012). Development and plasticity of the primary visual cortex. *Neuron*, *75*(2), 230-249. https://doi.org/10.1016/j.neuron.2012.06.009

Flavell, J. H. (1994). Cognitive development: Past, present, and future. In R. D. Parke, P. A. Ornstein, J. J. Rieser, & C. Zahn-Waxler (Eds.), *A century of developmental psychology* (pp. 569–587). American Psychological Association https://doi.org/10.1037/10155-020

Flom, R., Whipple, H., & Hyde, D. (2009). Infants' intermodal perception of canine (Canis familairis) facial expressions and vocalizations. *Developmental Psychology*, *45*(4), 1143. https://doi.org/10.1037/a0015367

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford University Press. https://doi.org/10.1093/0198236360.001.0001

Folland, N. A., Butler, B. E., Smith, N. A., & Trainor, L. J. (2012). Processing simultaneous auditory objects: Infants' ability to detect mistuning in harmonic complexes. *The Journal of the Acoustical Society of America*, *131*(1), 993-997. https://doi.org/10.1121/1.3651254

References

Folsom, R. C., & Wynne, M. K. (1987). Auditory brain stem responses from human adults and infants: Wave V tuning curves. *The Journal of the Acoustical Society of America*, *81*(2), 412-417. https://doi.org/10.1121/1.394906

Forssberg, H., Eliasson, A. C., Kinoshita, H., Westling, G., & Johansson, R. S. (1995). Development of human precision grip. *Experimental Brain Research*, *104*(2), 425-433.

Fox, R., & McDaniel, C. (1982). The perception of biological motion by human infants. *Science*, *218*(4571), 486-487. https://doi.org/10.1126/science.7123249

Fox, R., Aslin, R. N., Shea, S. L., & Dumais, S. T. (1980). Stereopsis in human infants. *Science, 207*(4428), 323–324. https://doi.org/10.1126/science.7350666

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience, 6*(4), 187–214. https://doi.org/10.1080/17588928.2015.1020053

Gaissert, N., & Wallraven, C. (2012). Categorizing natural objects: a comparison of the visual and the haptic modalities. *Experimental Brain Research*, *216*, 123-134. https://doi.org/10.1007/s00221-011-2916-4

Gandolfi, A., Horoupian, D. S., & De Teresa, R. M. (1981). Quantitative and cytometric analysis of the ventral cochlear nucleus in man. *Journal of the Neurological Sciences*, *50*(3), 443-455. https://doi.org/10.1016/0022-510X(81)90156-8

Gao, M., Turner, B. M., & Sloutsky, V. M. (2024). The role of attention in category representation. *Cognitive Science*, *48*(4), e13438. https://doi.org/10.1111/cogs.13438

Gattass, R., Soares, J. G., & Lima, B. (2017). Connectivity of the Pulvinar. In *The pulvinar thalamic nucleus of non-human primates: architectonic and functional*

subdivisions (pp. 19-29). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-70046-5_5

Gentaz, E., Baud-Bovy, G. & Luyat, M. (2008). The haptic perception of spatial orientations. *Experimental Brain Research,* 187, 331–348. https://doi.org/10.1007/s00221-008-1382-0

Gershkoff-Stowe, L., & Smith, L. B. (2004). Shape and the first hundred nouns. *Child Development*, *75*(4), 1098-1114. https://doi.org/10.1111/j.1467-8624.2004.00728.x

Ghim, H. R., & Eimas, P. D. (1988). Global and local processing by 3-and 4-month-old infants. *Attention*, *Perception & Psychophysics*, *43*(2), 165-171. https://doi.org/10.3758/BF03214194

Gibson, E. J., & Walker, A. S. (1984). Development of knowledge of visual-tactual affordances of substance. *Child Development*, 453-460. https://doi.org/10.2307/1129956

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., Paus, T., Evans, A. C., & Rapoport, J. L. (1999). Brain development during childhood and adolescence: A longitudinal MRI study. *Nature Neuroscience, 2*(10), 861–863. https://doi.org/10.1038/13158

Gliner, C. R. (1967). Tactual discrimination thresholds for shape and texture in young children. *Journal of Experimental Child Psychology*, *5*(4), 536-547. https://doi.org/10.1016/0022-0965(67)90048-3

Gogate, L. J., & Bahrick, L. E. (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of Experimental Child Psychology, 69*(2), 133–149. https://doi.org/10.1006/jecp.1998.2438

References

Gogate, L. J., & Bahrick, L. E. (2001). Intersensory redundancy and 7-month-old infants' memory for arbitrary syllable–object relations. *Infancy, 2*(2), 219–231. https://doi.org/10.1207/S15327078IN0202_7

Gogate, L. J., Bolzani, L. H., & Betancourt, E. A. (2006). Attention to maternal multimodal naming by 6-to 8-month-old infants and learning of word–object relations. *Infancy*, *9*(3), 259-288. https://doi.org/10.1207/s15327078in0903_1

Golarai, G., Ghahremani, D. G., Whitfield-Gabrieli, S., Reiss, A., Eberhardt, J. L., & Grill-Spector, K. (2007). Differential development of high-level visual cortex correlates with category-specific recognition memory. *Nature Neuroscience, 10*(4), 512–522. https://doi.org/10.1038/nn1865

Goldenberg, E. R., & Johnson, S. P. (2015). Category generalization in a new context: The role of visual attention. *Infant Behavior and Development*, *38*, 49-56. https://doi.org/10.1016/j.infbeh.2014.12.001

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49*, 585–612. https://doi.org/10.1146/annurev.psych.49.1.585

Goldstone, R.L. (1994) The role of similarity in categorization: providing a groundwork. *Cognition*, *52*, 125–157. https://doi.org/10.1016/0010-0277(94)90065-5

Gorga, M. P., Kaminski, J. R., Beauchaine, K. A., & Jesteadt, W. (1988). Auditory brainstem responses to tone bursts in normally hearing subjects. *Journal of Speech, Language, and Hearing Research*, *31*(1), 87-97. https://doi.org/10.1044/jshr.3101.87

Gori, M. (2015). Multisensory integration and calibration in children and adults with and without sensory and motor disabilities. *Multisensory Research*, *28*(1-2), 71-99. https://doi.org/10.1163/22134808-00002478

References

Gori, M., Del Viva, M., Sandini, G., & Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Current Biology*, *18*(9), 694-698. https://doi.org/10.1016/j.cub.2008.04.036

Gori, M., Giuliana, L., Sandini, G., & Burr, D. (2012). Visual size perception and haptic calibration during development. *Developmental Science*, *15*(6), 854-862. https://doi.org/10.1111/j.1467-7687.2012.01183.x

Gori, M., Sandini, G. and Burr, D. (2012). Development of visuo-auditory integration in space and time, *Frontiers in Integrative Neuroscience,* 6, 77. https://doi.org/10.3389/fnint.2012.00077

Gori, M., Sandini, G., Martinoli, C., & Burr, D. (2010). Poor haptic orientation discrimination in nonsighted children may reflect disruption of cross-sensory calibration. *Current Biology*, *20*(3), 223-225. https://doi.org/10.1016/j.cub.2009.11.069

Gori, M., Squeri, V., Sciutti, A., Masia, L., Sandini, G., & Konczak, J. (2012). Motor commands in children interfere with their haptic perception of objects. *Experimental Brain Research*, *223*(1), 149-157. https://doi.org/10.1007/s00221-012-3248-8

Gori, M., Vercillo, T., Sandini, G., & Burr, D. C. (2013). Children do not recalibrate motor-sensory temporal order after exposure to delayed sensory feedback. *Developmental Science*, 18(5), 703-712. https://doi.org/10.1111/desc.12247

Gougoux, F., Zatorre, R. J., Lassonde, M., Voss, P., & Lepore, F. (2005). A functional neuroimaging study of sound localization: Visual cortex activity predicts performance in early-blind individuals. *PLoS Biology, 3*(2), e27. https://doi.org/10.1371/journal.pbio.0030027

References

Graham, S. A., & Diesendruck, G. (2010). Fifteen-month-old infants attend to shape over other perceptual properties in an induction task. *Cognitive Development*, *25*(2), 111-123. https://doi.org/10.1016/j.cogdev.2009.06.002

Graham, S. A., & Poulin-Dubois, D. (1999). Infants' reliance on shape to generalize novel labels to animate and inanimate objects. *Journal of Child Language*, *26*(2), 295-320. https://doi.org/10.1017/S0305000999003815

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*(8), 357-364. https://doi.org/10.1016/j.tics.2010.05.004

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), *Cambridge handbook of computational psychology* (pp. 59–100). Cambridge University Press.

Hadad, B., Schwartz, S., Maurer, D., & Lewis, T. L. (2015). Motion perception: a review of developmental changes and the role of early visual experience. *Frontiers in Integrative Neuroscience*, *9*, 49. https://doi.org/10.3389/fnint.2015.00049

Haith, M. M. (2013). Future-oriented processes in infancy: The case of visual expectations. In *Visual Perception and Cognition in Infancy* (pp. 235-264). Psychology Press.

Hannon, E. E., & Johnson, S. P. (2005). Infants use meter to categorize rhythms and melodies: Implications for musical structure learning. *Cognitive Psychology*, *50*(4), 354-377. https://doi.org/10.1016/j.cogpsych.2004.09.003

Hannon, E. E., Schachner, A., & Nave-Blodgett, J. E. (2017). Babies know bad dancing when they see it: Older but not younger infants discriminate between synchronous

and asynchronous audiovisual musical displays. *Journal of Experimental Child Psychology*, *159*, 159-174. https://doi.org/10.1016/j.jecp.2017.01.006

Hartley, D. E., Wright, B. A., Hogan, S. C., & Moore, D. R. (2000). Age-related improvements in auditory backward and simultaneous masking in 6-to 10-year-old children. *Journal of Speech, Language, and Hearing Research*, *43*(6), 1402-1415. https://doi.org/10.1044/jslhr.4306.140

Hatwell, Y., Streri, A. & Gentaz, E. (2003). *Touching for Knowing: Cognitive psychology of haptic manual perception*. John Benjamins Publishing Company. https://doi.org/10.1075/aicr.53

Hayes, B. K., & Taplin, J. E. (1993). Developmental differences in the use of prototype and exemplar-specific information. *Journal of Experimental Child Psychology*, *55*(3), 329-352. https://doi.org/10.1006/jecp.1993.1019

Heikkilä, J., & Tiippana, K. (2015). Audiovisual semantic congruency during encoding enhances memory performance in school-aged children. *Quarterly Journal of Experimental Psychology, 68*(8), 1673–1690. https://doi.org/10.1027/1618-3169/a000279

Heikkilä, J., & Tiippana, K. (2016). School-aged children can benefit from audiovisual semantic congruency during memory encoding. *Experimental Brain Research*, *234*, 1199–1207. https://doi.org/10.1007/s00221-015-4341-6

Helbig, H. B., & Ernst, M. O. (2007). Optimal integration of shape information from vision and touch. *Experimental Brain Research, 179*(4), 595–606. https://doi.org/10.1007/s00221-006-0814-y

Hellman, R., Miśkiewicz, A., & Scharf, B. (1997). Loudness adaptation and excitation patterns: Effects of frequency and level. *The Journal of the Acoustical Society of America*, *101*(4), 2176-2185. https://doi.org/10.1121/1.418202

References

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

Hendrickson, A., Possin, D., Vajzovic, L., & Toth, C. A. (2012). Histologic development of the human fovea from midgestation to maturity. *American journal of ophthalmology*, *154*(5), 767-778. https://doi.org/10.1016/j.ajo.2012.05.007

Hendrickson, K., Love, T., Walenski, M., & Friend, M. (2019). The organization of words and environmental sounds in the second year: Behavioral and electrophysiological evidence. *Developmental science*, *22*(1), e12746. https://doi.org/10.1111/desc.12746

Hensch, T. K. (2005). Critical period mechanisms in developing visual cortex. *Current Topics in Developmental Biology*, *69*, 215-237. https://doi.org/10.1016/S0070-2153(05)69008-4

Heron, J., Whitaker, D., & McGraw, P. V. (2004). Sensory uncertainty governs the extent of audio-visual interaction. *Vision Research*, *44*(25), 2875-2884. https://doi.org/10.1016/j.visres.2004.07.001

Hillock, A. R., Powers, A. R., & Wallace, M. T. (2011). Binding of sights and sounds: Age-related changes in multisensory temporal processing. *Neuropsychologia, 49*(3), 461–467. https://doi.org/10.1016/j.neuropsychologia.2010.11.041

Hillock-Dunn, A., & Wallace, M. T. (2012). Developmental changes in the multisensory temporal binding window persist into adolescence. *Developmental Science, 15*(5), 688–696. https://doi.org/10.1111/j.1467-7687.2012.01171.x

Hoehl, S. (2016). The development of category specificity in infancy–What can we learn from electrophysiology? *Neuropsychologia*, *83*, 114-122. https://doi.org/10.1016/j.neuropsychologia.2015.08.021

References

Hofrichter, R., Siddiqui, H., Morrisey, M. N., & Rutherford, M. D. (2021). Early attention to animacy: Change-detection in 11-month-olds. *Evolutionary Psychology*, *19*(2). https://doi.org/10.1177/14747049211028220

Horton, M. S., & Markman, E. M. (1980). Developmental differences in the acquisition of basic and superordinate categories. *Child Development*, 708-719. https://doi.org/10.2307/1129456

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

Huttenlocher, P. R. (1999). Synaptogenesis in human cerebral cortex and the concept of critical periods. *The role of early experience in infant development*, 15-28.

Huyck, J. J., & Wright, B. A. (2011). Late maturation of auditory perceptual learning. *Developmental Science*, *14*(3), 614-621. https://doi.org/10.1111/j.1467-7687.2010.01009.x

Iordanescu, L., Grabowecky, M., & Suzuki, S. (2008). Characteristic sounds facilitate visual search. *Psychonomic Bulletin & Review, 15*(3), 548–554. https://doi.org/10.3758/PBR.15.3.548

Iordanescu, L., Guzman-Martinez, E., Grabowecky, M., & 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.72.7.1736

Jao, R. J., James, T. W., & Sathian, K. (2014). Multisensory convergence of visual and haptic object preference in lateral occipital complex. *Neuropsychologia, 61*, 124–134. https://doi.org/10.1016/j.neuropsychologia.2014.02.009

References

Jao, R. J., James, T. W., & Sathian, K. (2015). Crossmodal enhancement in the LOC for visuo-haptic object recognition depends on object familiarity. *Frontiers in Human Neuroscience, 9*, 513. https://doi.org/10.1016/j.neuropsychologia.2015.08.008

Jensen, J. K., & Neff, D. L. (1993). Development of basic auditory discrimination in preschool children. *Psychological Science*, *4*(2), 104-107. https://doi.org/10.1111/j.1467-9280.1993.tb00469.x

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*(2), 201-211. https://doi.org/10.3758/BF03212378

Johnson, C. E. (2000). Childrens' phoneme identification in reverberation and noise. *Journal of Speech, Language, and Hearing Research*, *43*(1), 144-157. https://doi.org/10.1044/jslhr.4301.144

Johnson, M. H. (2001). Functional brain development in humans. *Nature Reviews Neuroscience, 2*(7), 475–483. https://doi.org/10.1038/35081509

Johnson, M. H. (2005). Subcortical face processing. *Nature Reviews Neuroscience, 6*(10), 766–774. https://doi.org/10.1038/nrn1766

Johnson, M. H. (2011). Interactive specialization: a domain-general framework for human functional brain development?. *Developmental Cognitive Neuroscience*, *1*(1), 7-21. https://doi.org/10.1016/j.dcn.2010.07.003

Jordan, K. E., & Baker, J. (2011). Multisensory information boosts numerical matching abilities in young children. *Developmental Science*, *14*(2), 205-213. https://doi.org/10.1111/j.1467-7687.2010.00966.x

Jusczyk, P. W., & Aslin, R. N. (1995). Infants′ detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*(1), 1-23. https://doi.org/10.1006/cogp.1995.1010

Jusczyk, P. W., & Thompson, E. (1978). Perception of a phonetic contrast in multisyllabic utterances by 2-month-old infants. *Perception & Psychophysics*, *23*(2), 105-109. https://doi.org/10.3758/bf03208289

Kaganovich, N. (2017). Sensitivity to audiovisual temporal asynchrony in children with a history of specific language impairment and their peers with typical development: A replication and follow-up study. *Journal of Speech, Language, and Hearing Research*, *60*(8), 2259-2270. https://doi.org/10.1044/2017_JSLHR-L-16-03

Kalagher, H., & Jones, S. S. (2011a). Young children's haptic exploratory procedures. *Journal of Experimental Child Psychology*, *110*(4), 592-602. https://doi.org/10.1016/j.jecp.2011.06.007

Kalagher, H., & Jones, S. S. (2011b). Developmental change in young children's use of haptic information in a visual task: The role of hand movements. *Journal of Experimental Child Psychology*, *108*(2), 293-307. https://doi.org/10.1016/j.jecp.2010.09.004

Káldy, Z., & Kovács, I. (2003). Visual context integration is not fully developed in 4-year-old children. *Perception, 32*(6), 657–666. https://doi.org/10.1068/p3473

Kar, K., Kubilius, J., Schmidt, K., Issa, E. B., & DiCarlo, J. J. (2019). Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nature Neuroscience, 22*(6), 974–983. https://doi.org/10.1038/s41593-019-0392-5

Kaur, S., Espenhahn, S., Bell, T., Godfrey, K. J., Nwaroh, C., Giuffre, A., Beltrano, W., Yan, T., Stokoe, M. & Harris, A. D. (2022). Nonlinear age effects in tactile processing from early childhood to adulthood. *Brain and Behavior*, *12*(7), e2644. https://doi.org/10.1002/brb3.2644

References

Kay, M., & Wobbrock, J. (2021). ARTool: aligned rank transform for nonparametric factorial ANOVAs. *R package version 0.11*, *1*(10.5281).

Kellman, P. J., Arterberry, M. E., Damon, W., Lerner, R. M., Kuhn, D., & Siegler, R. S. (2006). Infant Visual Perception. *Handbook of Child Psychology, 2.*

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review Psychology*, *55*(1), 271-304.n https://doi.org/10.1146/annurev.psych.55.090902.142005

Kim, R. S., Seitz, A., & Shams, L. (2008). Benefits of stimulus congruency for multisensory facilitation of visual learning. *PLOS ONE, 3*(1), e1532. https://doi.org/10.1371/journal.pone.0001532

King, B. R., Oliveira, M. A., Contreras-Vidal, J. L., & Clark, J. E. (2012). Development of state estimation explains improvements in sensorimotor performance across childhood. *Journal of Neurophysiology*, *107*(11), 3040-3049. https://doi.org/10.1152/jn.00932.2011

Kirkham, N. Z., Rea, M., Osborne, T., White, H., & Mareschal, D. (2019). Do cues from multiple modalities support quicker learning in primary schoolchildren?. *Developmental Psychology*, *55*(10), 2048. https://doi.org/10.1037/dev0000778

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35-B42. https://doi.org/10.1016/S0010-0277(02)00004-5

Klatzky, R. L., Lederman, S. J., & Matula, D. E. (1993). Haptic exploration in the presence of vision. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(4), 726. https://doi.org/10.1037/0096-1523.19.4.726

References

Klatzky, R. L., Lederman, S. J., & Metzger, V. A. (1985). Identifying objects by touch: An "expert system". *Attention, Perception, & Psychophysics, 37*(4), 299–302. https://doi.org/10.3758/BF03211351

Knöpfel, T., Sweeney, Y., Radulescu, C. I., Zabouri, N., Doostdar, N., Clopath, C., & Barnes, S. J. (2019). Audio-visual experience strengthens multisensory assemblies in adult mouse visual cortex. *Nature Communications*, *10*(1), 5684. https://doi.org/10.1038/s41467-019-13607-2

Komar, G. F., Mieth, L., Buchner, A., & Bell, R. (2023). Animacy enhances recollection but not familiarity: Convergent evidence from the remember-know-guess paradigm and the process-dissociation procedure. *Memory & Cognition*, *51*(1), 143-159. https://doi.org/10.3758/s13421-022-01339-6

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS ONE, 2*(9), e943. https://doi.org/10.1371/journal.pone.0000943

Kosakowski, H. L., Cohen, M. A., Herrera, L., Nichoson, I., Kanwisher, N., & Saxe, R. (2024). Cortical face-selective responses emerge early in human infancy. *eneuro*, *11*(7). https://doi.org/10.1523/ENEURO.0117-24.2024

Kosakowski, H. L., Cohen, M. A., Takahashi, A., Keil, B., Kanwisher, N., & Saxe, R. (2022). Selective responses to faces, scenes, and bodies in the ventral visual pathway of infants. *Current Biology*, *32*(2), 265-274. https://doi.org/10.1016/j.cub.2021.10.064

Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science, 293*(5534), 1506–1509. https://doi.org/10.1126/science.1061133

References

Kourtzi, Z., Erb, M., Grodd, W., & Bülthoff, H. H. (2003). Representation of the perceived 3D object shape in the human lateral occipital complex. *Cerebral Cortex, 13*(9), 911–920. https://doi.org/10.1093/cercor/13.9.911

Kovács, I. (2000). Human development of perceptual organization. *Vision Research, 40*(10–12), 1301–1310. https://doi.org/10.1016/S0042-6989(00)00055-9

Kraebel, K. S., & Gerhardstein, P. C. (2006). Three-month-old infants' object recognition across changes in viewpoint using an operant learning procedure. *Infant Behavior and Development*, *29*(1), 11-23. https://doi.org/10.1016/j.infbeh.2005.10.002

Krishna, O., Yamasaki, T., Helo, A., Pia, R., & Aizawa, K. (2017). Developmental changes in ambient and focal visual processing strategies. *Electronic Imaging*, *29*, 224-229. https://doi.org/10.2352/ISSN.2470-1173.2017.14.HVEI-148

Krishnan, A., Gandour, J.T. (2017). Shaping Brainstem Representation of Pitch-Relevant Information by Language Experience. In: Kraus, N., Anderson, S., White-Schwoch, T., Fay, R., Popper, A. (eds) The Frequency-Following Response. Springer Handbook of Auditory Research, vol 61. Springer, Cham. https://doi.org/10.1007/978-3-319-47944-6_3

Krishnan, S., Leech, R., Aydelott, J., & Dick, F. (2013). School-age children's environmental object identification in natural auditory scenes: Effects of masking and contextual congruence. *Hearing Research*, *300*, 46-55. https://doi.org/10.1016/j.heares.2013.03.003

Krumholz, A., Felix, J. K., Goldstein, P. J., & McKenzie, E. (1985). Maturation of the brain-stem auditory evoked potential in premature infants. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *62*(2), 124-134. https://doi.org/10.1016/0168-5597(85)90024-3

References

Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, *45*(6), 812-863. https://doi.org/10.1006/jmps.2000.1354

Kuhlmeier, V. A., Troje, N. F., & Lee, V. (2010). Young infants detect the direction of biological motion in point-light displays. *Infancy*, *15*(1), 83-93. https://doi.org/10.1111/j.1532-7078.2009.00003.x

Lacey, S., & Sathian, K. (2014). Visuo-haptic multisensory object recognition, categorization, and representation. *Frontiers in Psychology, 5*, 730. https://doi.org/10.3389/fpsyg.2014.00730

Lacey, S., Flueckiger, P., Stilla, R., Lava, M., & Sathian, K. (2010). Object familiarity modulates the relationship between visual object imagery and haptic shape perception. *Neuroimage*, *49*(3), 1977-1990. https://doi.org/10.1016/j.neuroimage.2009.10.081

Lacey, S., Peters, A., & Sathian, K. (2007). Cross-modal object recognition is viewpoint-independent. *PLoS ONE, 2*(9), e890. https://doi.org/10.1371/journal.pone.0000890

Landau, B., Smith, L. B., & Jones, S. (1988a). The importance of shape in early lexical learning. *Cognitive Development, 3*(3), 299–321. https://doi.org/10.1016/0885-2014(88)90014-7

Landau, B., Smith, L. B., & Jones, S. (1998b). Object shape, object function, and object name. *Journal of Memory and Language*, *38*(1), 1-27. https://doi.org/10.1006/jmla.1997.2533

Landau, B., Smith, L., & Jones, S. (1998c). Object perception and object naming in early development. *Trends in Cognitive Sciences*, *2*(1), 19-24. https://doi.org/10.1016/S1364-6613(97)01111-X

References

Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research, 158*(4), 405–414. https://doi.org/10.1007/s00221-004-1913-2

Lavigne-Rebillard, M., & Pujol, R. (1988). Hair cell innervation in the fetal human cochlea. *Acta Oto-Laryngologica*, *105*(5-6), 398-402. https://doi.org/10.3109/00016488809119492

Lebenberg, J., Mangin, J. F., Thirion, B., Poupon, C., Hertz-Pannier, L., Leroy, F., Adibpour, P. , Dehaene-Lambertz, G., & Dubois, J. (2019). Mapping the asynchrony of cortical maturation in the infant brain: A MRI multi-parametric clustering approach. *Neuroimage*, *185*, 641-653. https://doi.org/10.1016/j.neuroimage.2018.07.022

Lecanuet, J. P., & Schaal, B. (1996). Fetal sensory competencies. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, *68*, 1-23. https://doi.org/10.1016/0301-2115(96)02509-2

Lecours, A. R. (1975). Myelogenetic correlates of the development of speech and language. In *Foundations of language development* (pp. 121-135). Academic Press. https://doi.org/10.1016/B978-0-12-443701-2.50017-2

Lederman, S. J., & Klatzky, R. L. (1987). Hand movements: A window into haptic object recognition. *Cognitive Psychology, 19*(3), 342–368. https://doi.org/10.1016/0010-0285(87)90008-9

Lederman, S. J., & Klatzky, R. L. (2009). Haptic perception: A tutorial. *Attention, Perception, & Psychophysics*, *71*(7), 1439-1459. https://doi.org/10.3758/APP.71.7.1439

References

Lee, M. D., & Vanpaemel, W. (2008). Exemplars, prototypes, similarities, and rules in category representation: An example of hierarchical Bayesian analysis. *Cognitive Science*, *32*(8), 1403-1424. https://doi.org/10.1080/03640210802073697

Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press. https://doi.org/10.1017/CBO9781139087759

Lehmann, S., & Murray, M. M. (2005). The role of multisensory memories in unisensory object discrimination. *Cognitive Brain Research*, *24*(2), 326-334. https://doi.org/10.1016/j.cogbrainres.2005.02.005

Leibold, L. J., & Werner, L. A. (2002). Relationship between intensity and reaction time in normal-hearing infants and adults. *Ear and Hearing*, *23*(2), 92-97. https://doi.org/10.1097/00003446-200204000-00002

Lejeune, F., Audeoud, F., Marcus, L., Streri, A., Debillon, T., & Gentaz, E. (2010). The manual habituation and discrimination of shapes in preterm human infants from 33 to 34+ 6 post-conceptional age. *PLoS One*, *5*(2), e9108. https://doi.org/10.1371/journal.pone.0009108

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality?. *Cognition*, *25*(3), 265-288. https://doi.org/10.1016/S0010-0277(87)80006-9

Lewis, J. W., Talkington, W. J., Walker, N. A., Spirou, G. A., Jajosky, A., Frum, C., & Brefczynski-Lewis, J. A. (2009). Human cortical organization for processing vocalizations indicates representation of harmonic structure as a signal attribute. *Journal of Neuroscience, 29*(7), 2283–2296. https://doi.org/10.1523/JNEUROSCI.4145-08.2009

Lewkowicz, D. J. (1988a). Sensory dominance in infants: I. Six-month-old infants' response to auditory-visual compounds. *Developmental Psychology*, *24*(2), 155. http://dx.doi.org/10.1037/0012-1649.24.2.155

References

Lewkowicz, D. J. (1988b). Sensory dominance in infants: II. Ten-month-old infants' response to auditory-visual compounds. *Developmental Psychology*, *24*(2), 172. https://doi.org/10.1037/0012-1649.24.2.172

Lewkowicz, D. J. (1992). Infants' responsiveness to the auditory and visual attributes of a sounding/moving stimulus. *Perception & Psychophysics, 52*(5), 519–528. https://doi.org/10.3758/BF03206713

Lewkowicz, D. J. (1996). Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception and Performance, 22*(5), 1094–1106. https://doi.org/10.1037/0096-1523.22.5.1094

Lewkowicz, D. J. (2000). The development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychological Bulletin, 126*(2), 281–308. https://doi.org/10.1037/0033-2909.126.2.281

Lewkowicz, D. J. (2010). Infant perception of audio-visual speech synchrony. *Developmental Psychology, 46*(1), 66–77. https://doi.org/10.1037/a0015579

Lewkowicz, D. J. (2014). Early experience and multisensory perceptual narrowing. *Developmental Psychobiology, 56*(2), 292–315. https://doi.org/10.1002/dev.21197

Lewkowicz, D. J., & Flom, R. (2014). The audiovisual temporal binding window narrows in early childhood. *Child Development, 85*(2), 685–694. https://doi.org/10.1111/cdev.12142

Lewkowicz, D. J., & Ghazanfar, A. A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in Cognitive Sciences, 13*(11), 470–478. https://doi.org/10.1016/j.tics.2009.08.004

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. Proceedings of the National

Academy of Sciences, 109(5), 1431-1436. https://doi.org/10.1073/pnas.1114783109

Lewkowicz, D. J., & Kraebel, K. S. (2004). The Value of Multisensory Redundancy in the Development of Intersensory Perception. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 655–678). Boston Review. https://doi.org/10.7551/mitpress/3422.003.0049

Lewkowicz, D. J., & Lickliter, R. (1998). The detection of intermodal rate and synchrony relations in 6-and 8-month-old infants. *Infant Behavior and Development*, *21*, 537.

Li, J., & Deng, S. W. (2023). Facilitation and interference effects of the multisensory context on learning: a systematic review and meta-analysis. *Psychological Research*, *87*(5), 1334-1352. https://doi.org/10.1007/s00426-022-01733-4

Li, Q., Wu, Q., Yu, Y., Wu, F., Takahashi, S., Ejima, Y., Yang, J., & Wu, J. (2020). Semantic congruency modulates the effect of attentional load on the audiovisual integration of animate images and sounds. *i-Perception*, *11*(6), https://doi.org/10.1177/2041669520981096

Lickliter, R. (1990). The role of sensory stimulation in perinatal development: Insights from comparative research for the care of the high-risk infant. *Journal of Developmental and Behavioral Pediatrics, 11*(4), 226–232. https://doi.org/10.1097/00004703-200012000-00007

Lickliter, R., & Bahrick, L. E. (2004). Perceptual Development and the Origins of Multisensory Responsiveness. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 643–654). Boston Review. https://doi.org/10.7551/mitpress/3422.003.0048

Litovsky, R. (2015). Development of the auditory system. Handbook of clinical neurology, 129, 55-72. https://doi.org/10.1016/B978-0-444-62630-1.00003-2

References

Litovsky, R. Y. (1997). Developmental changes in the precedence effect: Estimates of minimum audible angle. *The Journal of the Acoustical Society of America*, *102*(3), 1739-1745. https://doi.org/10.1121/1.420106

Litovsky, R. Y., Ashmead, D. H., Gilkey, R., & Anderson, T. (1997). Development of binaural and spatial hearing in infants and children. *Binaural and spatial hearing in real and virtual environments*, 571-592. https://doi.org/10.1007/978-1-4614-1421-6_6

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, *5*(5), 552-563. https://doi.org/10.1016/S0960-9822(95)00108-4

López, A., Gelman, S. A., Gutheil, G., & Smith, E. E. (1992). The development of category-based induction. *Child Development*, *63*(5), 1070-1090. https://doi.org/10.1111/j.1467-8624.1992.tb01681.x

Lu, K., Xu, Y., Yin, P., Oxenham, A. J., Fritz, J. B., & Shamma, S. A. (2017). Temporal coherence structure rapidly shapes neuronal interactions. *Nature Communications*, *8*(1), 13900. https://doi.org/10.1038/ncomms13900

Luna, B., Velanova, K., & Geier, C. F. (2008). Development of eye-movement control. *Brain and Cognition*, *68*(3), 293-308. https://doi.org/10.1016/j.bandc.2008.08.019

Lunghi, M., & Di Giorgio, E. (2024). I like the way you move: how animate motion affects visual attention in early human infancy. *Frontiers in Neuroscience*, *18*, 1459550. https://doi.org/10.3389/fnins.2024.1459550

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology, 3*, 54. https://doi.org/10.3389/fpsyg.2012.00054

References

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience, 9*(11), 1432–1438. https://doi.org/10.1038/nn1790

Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioral Processes, 66*(3), 309–332. https://doi.org/10.1016/j.beproc.2004.03.011

Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(4), 650–662. https://doi.org/10.1037/0278-7393.29.4.650

Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2005). Delayed feedback disrupts the procedural-learning system but not the hypothesis-testing system in perceptual category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 100–107. https://doi.org/10.1037/0278-7393.31.1.100

Mandler, J. M. (2000). Perceptual and conceptual processes in infancy. *Journal of Cognition and Development, 1*(1), 3–36. https://doi.org/10.1207/S15327647JCD0101N_2

Mandler, J. M. (2008). On the birth and growth of concepts. *Philosophical Psychology*, *21*(2), 207-230. https://doi.org/10.1080/09515080801980179

Mandler, J. M., & Bauer, P. J. (1988). The cradle of categorization: Is the basic level basic?. *Cognitive Development*, *3*(3), 247-264. https://doi.org/10.1016/0885-2014(88)90011-1

Mandler, J. M., Bauer, P. J., & McDonough, L. (1991). Separating the sheep from the goats: Differentiating global categories. *Cognitive Psychology*, *23*(2), 263-298. https://doi.org/10.1016/0010-0285(91)90011-C

References

Mareschal, D., & Quinn, P. C. (2001). Categorization in infancy. *Trends in Cognitive Sciences*, *5*(10), 443-450. https://doi.org/10.1016/S1364-6613(00)01752-6

Mareschal, D., & Thomas, M. S. (2007). Computational modeling in developmental psychology. *IEEE Transactions on Evolutionary Computation*, *11*(2), 137-150. https://doi.org/10.1109/TEVC.2006.890232

Mareschal, D., French, R. M., & Quinn, P. C. (2000). A connectionist account of asymmetric category learning in early infancy. *Developmental Psychology*, *36*(5), 635. https://doi.org/10.1037/0012-1649.36.5.635

Mareschal, D., Powell, D., & Volein, A. (2003). Basic-level category discriminations by 7-and 9-month-olds in an object examination task. *Journal of Experimental Child Psychology*, *86*(2), 87-107. https://doi.org/10.1016/S0022-0965(03)00107-3

Mareschal, D., Powell, D., Westermann, G., & Volein, A. (2005). Evidence of rapid correlation-based perceptual category learning by 4-month-olds. *Infant and Child Development: An International Journal of Research and Practice*, *14*(5), 445-457. https://doi.org/10.1002/icd.415

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. MIT Press. https://doi.org/10.1017/S030500090001134X

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. Freeman. https://doi.org/10.7551/mitpress/9780262514620.001.0001

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society B: Biological Sciences, 200*(1140), 269–294. https://doi.org/10.1098/rspb.1978.0020

References

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature, 379*(6566), 649–652. https://doi.org/10.1038/379649a0

Mash, C. (2007). Object representation in infants' coordination of manipulative force. *Infancy*, *12*(3), 329-341. https://doi.org/10.1111/j.1532-7078.2007.tb00246.x

Mash, C., Arterberry, M. E., & Bornstein, M. H. (2007). Mechanisms of visual object recognition in infancy: Five-month-olds generalize beyond the view provided. *Developmental Science, 10*(3), 347–358. https://doi.org/10.1111/j.1532-7078.2007.tb00232.x

Matatyaho, D. J., & Gogate, L. J. (2008). Type of maternal object motion during synchronous naming predicts preverbal infants' learning of word–object relations. *Infancy*, *13*(2), 172-184. https://doi.org/10.1080/15250000701795655

Mateeff, S., Hohnsbein, J., & Noack, T. (1985). Dynamic visual capture: Apparent auditory motion induced by a moving visual target. *Perception, 14*(6), 721–727. https://doi.org/10.1068/p140721

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., Wallace, M. T., & Murray, M. M. (2017). A multisensory perspective on object memory. *Neuropsychologia*, *105*, 243-252. https://doi.org/10.1016/j.neuropsychologia.2017.04.008

References

Maxon, A. B., & Hochberg, I. (1982). Development of psychoacoustic behavior: Sensitivity and discrimination. *Ear and Hearing*, *3*(6), 301-308. https://doi.org/10.1097/00003446-198211000-00003

Medin, D. L., & Coley, J. D. (1998). Concepts and categorization. *Perception and cognition at century's end: Handbook of Perception and Cognition*, 403-439. https://doi.org/10.1016/B978-012301160-2/50015-0

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207. https://doi.org/10.1037/0033-295X.85.3.207

Meltzoff, A. N., & Borton, R. W. (1979). Intermodal matching by human neonates. *Nature, 282*(5737), 403–404. https://doi.org/10.1038/282403a0

Mercier, M. R., & Cappe, C. (2020). The interplay between multisensory integration and perceptual decision making. *NeuroImage, 222*, 116970. https://doi.org/10.1016/j.neuroimage.2020.116970

Millar, S. (1971). Visual and haptic cue utilization by preschool children: The recognition of visual and haptic stimuli presented separately and together. *Journal of Experimental Child Psychology*, *12*(1), 88–94. https://doi.org/10.1016/0022-0965(71)90019-1

Miller, C. L. (1983). Developmental changes in male/female voice classification by infants. *Infant Behavior and Development*, *6*(2-3), 313-330. https://doi.org/10.1016/S0163-6383(83)80040-X

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*(1), 167-202. https://doi.org/10.1146/annurev.neuro.24.1.167

References

Minda, J. P., & Miles, S. J. (2010). The influence of verbal and nonverbal processing on category learning. *Psychology of learning and motivation*, *52*, 117-162. https://doi.org/10.1016/S0079-7421(10)52003-6

Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus distortion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(3), 775–799. https://doi.org/10.1037/0278-7393.27.3.775

Minda, J. P., & Smith, J. D. (2011). Prototype models of categorization: Basic formulation, predictions, and limitations. In N. A. Stanton et al. (Eds.), Formal approaches in categorization. https://doi.org/10.1017/CBO9780511921322.003

Minda, J. P., Desroches, A. S., & Church, B. A. (2008). Learning rule-described and non-rule-described categories: a comparison of children and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1518. https://doi.org/10.1037/a0013355

Misceo, G. F., Hershberger, W. A., & Mancini, R. L. (1999). Haptic estimates of discordant visual—haptic size vary developmentally. *Perception & Psychophysics*, *61*(6), 608-614. https://doi.org/10.3758/BF03205533

Molholm, S., Murphy, J. W., Bates, J., Ridgway, E. M., & Foxe, J. J. (2020). Multisensory audiovisual processing in children with a sensory processing disorder (I): behavioral and electrophysiological indices under speeded response conditions. *Frontiers in Integrative Neuroscience*, *14*, 4. https://doi.org/10.3389/fnint.2020.00004

Molina, M., & Jouen, F. (1998). Modulation of the palmar grasp behavior in neonates according to texture property. *Infant Behavior and Development*, *21*(4), 659-666. https://doi.org/10.1016/S0163-6383(98)90036-4

References

Moore, C. J., & Price, C. J. (1999). A functional neuroimaging study of the variables that generate category-specific object processing differences. *Brain, 122*(5), 943–962. https://doi.org/10.1093/brain/122.5.943

Moore, D. R. (2012). Listening difficulties in children: Bottom-up and top-down contributions. *Journal of Communication Disorders, 45*(6), 411–418. https://doi.org/10.1016/j.jcomdis.2012.06.006

Moore, D. R., Ferguson, M. A., Halliday, L. F., & Riley, A. (2008). Frequency discrimination in children: Perception, learning and attention. *Hearing Research*, *238*(1-2), 147-154. https://doi.org/10.1016/j.heares.2007.11.013

Moore, J. K., & Linthicum Jr, F. H. (2001). Myelination of the human auditory nerve: different time courses for schwann cell and glial myelin. *Annals of Otology, Rhinology & Laryngology*, *110*(7), 655-661. https://doi.org/10.1177/000348940111000711

Moore, J. K., & Linthicum Jr, F. H. (2007). The human auditory system: a timeline of development. *International Journal of Audiology*, *46*(9), 460-478. https://doi.org/10.1080/14992020701383019

Moore, J. K., Guan, Y. L., & Shi, S. R. (1998). MAP2 expression in developing dendrites of human brainstem auditory neurons. *Journal of Chemical Neuroanatomy*, *16*(1), 1-15. https://doi.org/10.1016/S0891-0618(98)00057-X

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments: Special issue on interactive learning environments: Contemporary issues and trends. *Educational Psychology Review*, *19*, 309-326. https://doi.org/10.1007/s10648-007-9047-2

References

Morrongiello, B. A., & Rocca, P. T. (1987). Infants' localization of sounds in the horizontal plane: Effects of auditory and visual cues. *Child Development*, 918-927. https://doi.org/10.2307/1130532

Morrongiello, B. A., & Trehub, S. E. (1987). Age-related changes in auditory temporal perception. *Journal of Experimental Child Psychology*, *44*(3), 413-426. https://doi.org/10.1016/0022-0965(87)90043-9

Morrongiello, B. A., Fenwick, K. D., Hillier, L., & Chance, G. (1994). Sound localization in newborn human infants. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, *27*(8), 519-538. https://doi.org/10.1002/dev.420270805

Morrongiello, B. A., Humphrey, G. K., Timney, B., Choi, J., & Rocca, P. T. (1994). Tactual object exploration and recognition in blind and sighted children. *Perception, 23*(7), 833–848. https://doi.org/10.1068/p230833

Motoki, K., Marks, L. E., & Velasco, C. (2023). Reflections on cross-modal correspondences: Current understanding and issues for future research. *Multisensory Research*, *37*(1), 1-23. https://doi.org/10.1163/22134808-bja10114

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Percie du Sert, N., ... & Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 0021. https://doi.org/10.1038/s41562-016-0021

Murphy, G. L. (2002). *The big book of concepts.* Boston Review. https://doi.org/10.7551/mitpress/1602.001.0001

Murphy, G. L., & Wisniewski, E. J. (1989). Categorizing objects in isolation and in scenes: what a superordinate is good for. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, *15*(4), 572. https://doi.org/10.1037/0278-7393.15.4.572

Murray, C. A., & Shams, L. (2023). Crossmodal interactions in human learning and memory. *Frontiers in Human Neuroscience*, *17*, 1181760. https://doi.org/10.3389/fnhum.2023.1181760

Murray, M. M., & Wallace, M. T. (2011). *The neural bases of multisensory processes*. CRC press. https://doi.org/10.1201/9781439812174

Murray, M. M., Lewkowicz, D. J., Amedi, A., & Wallace, M. T. (2016). Multisensory processes: a balancing act across the lifespan. *Trends in Neurosciences*, *39*(8), 567-579. https://doi.org/10.1016/j.tins.2016.05.003

Murray, M. M., Thelen, A., Thut, G., Romei, V., Martuzzi, R., & Matusz, P. J. (2016). The multisensory function of the human primary visual cortex. *Neuropsychologia, 83*, 161–169. https://doi.org/10.1016/j.neuropsychologia.2015.08.011

Nacke, L. E., & Deterding, S. (2017). The maturing of gamification research. *Computers in Human Behavior*, *71*, 450-454.https://doi.org/10.1016/j.chb.2016.11.062

Nairne, J. S., VanArsdall, J. E., Pandeirada, J. N., Cogdill, M., & LeBreton, J. M. (2013). Adaptive memory: The mnemonic value of animacy. *Psychological Science*, *24*(10), 2099-2105. https://doi.org/10.1177/0956797613480803

Nara, T., Goto, N., Hamano, S. I., & Okada, A. (1996). Morphometric development of the human fetal auditory system: inferior collicular nucleus. *Brain and Development*, *18*(1), 35-39. https://doi.org/10.1016/0387-7604(95)00089-5

Nardini, M., Bales, J., & Mareschal, D. (2016). Integration of audio-visual information for spatial decisions in children and adults. *Developmental Science*, *19*(5), 803-816. https://doi.org/10.1111/desc.12327

References

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of cue integration in human navigation. *Current Biology, 18*(9), 689–693. https://doi.org/10.1016/j.cub.2008.04.021

Nava, E., & Pavani, F. (2013). Changes in sensory dominance during childhood: Converging evidence from the Colavita effect and the sound-induced flash illusion. *Child Development*, *84*(2), 604-616. https://doi.org/10.1111/j.1467-8624.2012.01856.x

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9*(3), 353–383. https://doi.org/10.1016/0010-0285(77)90012-3

Needham, A., Dueker, G., & Lockhead, G. (2005). Infants' formation and use of categories to segregate objects. *Cognition*, *94*(3), 215-240. https://doi.org/10.1016/j.cognition.2004.02.002

Negen, J., & Nardini, M. (2015). Four-year-olds use a mixture of spatial reference frames. *PloS one*, *10*(7), e0131984. https://doi.org/10.1371/journal.pone.0131984

Negen, J., Chere, B., Bird, L. A., Taylor, E., Roome, H. E., Keenaghan, S., Thaler, L., & Nardini, M. (2019). Sensory cue combination in children under 10 years of age. *Cognition*, *193*, 104014. https://doi.org/10.1016/j.cognition.2019.104014

Neijzen, C. M., de Wit, F. M., Hettinga, Y. M., de Boer, J. H., van Genderen, M. M., & de Wit, G. C. (2025). Reference values for the Teller Acuity Cards II (TAC II) in infants and preverbal children, a meta-analysis. *Acta Ophthalmologica*, *103*(4), 479-485. https://doi.org/10.1111/aos.17447

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

Nelson, K. (1988). Where do taxonomic categories come from? *Human Development*, *31*(1), 3-10. https://doi.org/10.1159/000273198

Newcombe, N., Rogoff, B., & Kagan, J. (1977). Developmental changes in recognition memory for pictures of objects and scenes. *Developmental Psychology*, *13*(4), 337. https://doi.org/10.1037/0012-1649.13.4.337

Newell, F. N. (2004). Cross-modal object recognition. In Calvert, G., Spence, C., & Stein, B. E. (Eds.), *The handbook of multisensory processes*. (pp. 123-139). MIT press. https://doi.org/10.7551/mitpress/3422.003.0011

Newell, F. N., Bülthoff, H. H., & Ernst, M. O. (2003). Cross-modal perception of actively explored objects. In S. O'Modhrain (Ed*.), Eurohaptics 2003 Conference Proceedings* (pp. 291–299). Dublin, ROI: Trinity College and Media Lab Europe. Retrieved from: https://hdl.handle.net/11858/00-001M-0000-0013-DC1C-6

Newell, F. N., Ernst, M. O., Tjan, B. S., & Bülthoff, H. H. (2001). Viewpoint dependence in visual and haptic object recognition. *Psychological Science, 12*(1), 37–42. https://doi.org/10.1111/1467-9280.00307

Newell, F. N., McKenna, E., Seveso, M. A., Devine, I., Alahmad, F., Hirst, R. J., & O'Dowd, A. (2023). Multisensory perception constrains the formation of object categories: A review of evidence from sensory-driven and predictive processes on categorical decisions. *Philosophical Transactions of the Royal Society B*, *378*(1886), 20220342. https://doi.org/10.1098/rstb.2022.0342

Newman, R. S., & Jusczyk, P. W. (1996). The cocktail party effect in infants. *Perception & Psychophysics*, *58*(8), 1145-1156. https://doi.org/10.3758/BF03207548

References

Nguyen, L., Silva, A. E., Poppe, T., Leung, M., Alsweiler, J. M., Black, J., Harding., J. E., & Yiang, Y. (2025). Early Nutrition is Associated With Global Motion Perception and V5 Function in 7-Year-Old Children Born Very Preterm. *Human Brain Mapping*, *46*(11), e70298. https://doi.org/10.1002/hbm.70298

Noppeney, U. (2021). Perceptual inference, learning, and attention in a multisensory world. *Annual Review of Neuroscience*, *44*(1), 449-473. https://doi.org/10.1146/annurev-neuro-100120-085519

Norcia, A. M., Tyler, C. W., & Hamer, R. D. (1990). Development of contrast sensitivity in the human infant. *Vision Research*, *30*(10), 1475-1486. https://doi.org/10.1016/0042-6989(90)90028-J

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D. Chin, G., ... & Yarkoni, T. (2015). Promoting an open research culture: Author guidelines for journals could help to promote transparency, openness, and reproducibility. *Science*. *348*(6242), 1422-1425. https://doi.org/10.1126/science.aab2374

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600-2606. https://doi.org/10.1073/pnas.1708274114

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39. https://doi.org/10.1037//0096-3445.115.1.39

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*(2), 266. https://doi.org/10.1037/0033-295x.104.2.266

References

O'Doherty, C., Dineen, Á. T., Truzzi, A., King, G., Zaadnoordijk, L., Harrison, K., D'Arcy, E. L, White, J., Caldinelli. C., Holloway, T., Kravchenko, A., Diedrichsen, J., Tarrant, A., Byrne, A. T., Foran, A., Mollow, E. J, & Cusack, R. (2026). Infants have rich visual categories in ventrotemporal cortex at 2 months of age. *Nature Neuroscience*, *29*, 693-702. https://doi.org/10.1038/s41593-025-02187-8

Oakes, L. M., & Rakison, D. H. (2003). Issues in the early development of concepts and categories: An introduction. In D. H. Rakison & L. M. Oakes (Eds.), *Early category and concept development: Making sense of the blooming, buzzing confusion* (pp. 3–23). Oxford University Press.

Oakes, L. M., & Spalding, T. L. (1997). The role of exemplar distribution in infants' differentiation of categories. *Infant Behavior and Development*, *20*(4), 457-475. https://doi.org/10.1016/S0163-6383(97)90036-9

Oehlschlaeger, S., & Vo, M. L. H. (2020). Development of scene knowledge: Evidence from explicit and implicit scene knowledge measures. *Journal of Experimental Child Psychology*, *194*, 104782. https://doi.org/10.1016/j.jecp.2019.104782

Ofen, N., Chai, X. J., Schuil, K. D., Whitfield-Gabrieli, S., & Gabrieli, J. D. (2012). The development of brain systems associated with successful memory retrieval of scenes. *Journal of Neuroscience*, *32*(29), 10012-10020. https://doi.org/10.1523/JNEUROSCI.1082-11.2012

Olsho, L. W. (1984). Infant frequency discrimination. *Infant Behavior and Development*, *7*(1), 27-35. https://doi.org/10.1016/S0163-6383(84)80020-X

Olsho, L. W., Koch, E. G., & Carter, E. A. (1988). Nonsensory factors in infant frequency discrimination. *Infant Behavior and Development*, *11*(2), 205-222. https://doi.org/10.1016/S0163-6383(88)80006-7

References

Olsho, L. W., Koch, E. G., & Halpin, C. F. (1987). Level and age effects in infant frequency discrimination. *The Journal of the Acoustical Society of America*, *82*(2), 454-464. https://doi.org/10.1121/1.395446

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/1126/science.aac4716

Opfer, J. E., & Gelman, S. A. (2011). Development of the animate-inanimate distinction. *The Wiley-Blackwell Handbook of Childhood Cognitive Development*, *2*, 213-238. https://doi.org/10.1002/9781444325485

Otsuka, Y., & Yamaguchi, M. K. (2003). Infants' perception of illusory contours in static and moving figures. *Journal of Experimental Child Psychology*, *86*(3), 244-251. https://doi.org/10.1016/S0022-0965(03)00126-7

Overvliet, K. E., Postma, A., & Röder, B. (2024). Child development and the role of visual experience in the use of spatial and non-spatial features in haptic object perception. *Journal of Experimental Child Psychology*, *242*, 105885. https://doi.org/10.1016/j.jecp.2024.105885

Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, *3*(5), 519-526. https://doi.org/10.3758/BF03197524

Palmeri, T. J., & Nosofsky, R. M. (2001). Central tendencies, extreme points, and prototype enhancement effects in ill-defined perceptual categorization. *Quarterly Journal of Experimental Psychology, 54A*(1), 197–235. https://doi.org/10.1080/02724980042000084

Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in macaque area V4. *Journal of Neurophysiology, 82*(5), 2490–2502. https://doi.org/10.1152/jn.1999.82.5.2490

References

Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area V4. *Nature Neuroscience, 5*(12), 1332–1338. https://doi.org/10.1038/972

Patterson, M. L., & Werker, J. F. (2002). Infants' ability to match dynamic phonetic and gender information in the face and voice. *Journal of Experimental Child Psychology*, *81*(1), 93-115. https://doi.org/10.1006/jecp.2001.2644

Pauen, S. (2002). Evidence for knowledge-based category discrimination in infancy. *Child. Development.* 73, 1016–1033. https://doi.org/10.1111/1467-8624.00454

Paulus, M., & Hauf, P. (2011). Infants' use of material properties to guide their actions with differently weighted objects. *Infant and Child Development*, *20*(4), 423-436. https://doi.org/10.1002/icd.704

Paus, T. (2005). Mapping brain maturation and cognitive development during adolescence. *Trends in Cognitive Sciences*, *9*(2), 60-68. https://doi.org/10.1016/j.tics.2004.12.008

Paus, T., Collins, D. L., Evans, A. C., Leonard, G., Pike, B., & Zijdenbos, A. (2001). Maturation of white matter in the human brain: A review of magnetic resonance studies. *Brain Research Bulletin, 54*(3), 255–266. https://doi.org/10.1016/S0361-9230(00)00434-2

Pavlova, M., Krägeloh-Mann, I., Sokolov, A., & Birbaumer, N. (2001). Recognition of point-light biological motion displays by young children. *Perception*, *30*(8), 925-933. https://doi.org/10.1068/p3157

Peebles, D., & Cooper, R. P. (2015). Thirty years after Marr's Vision: Levels of analysis in cognitive science. *Topics in Cognitive Science, 7*(2), 187–190. https://doi.org/10.1068/p3157

References

Peng, A., Kirkham, N. Z., & Mareschal, D. (2018). Task switching costs in preschool children and adults. *Journal of Experimental Child Psychology*, *172*, 59-72. https://doi.org/10.1016/j.jecp.2018.01.019

Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2011). The shape of the vocabulary predicts the shape of the bias. *Frontiers in Psychology, 2*, 345. https://doi.org/10.3389/fpsyg.2011.00345

Petrini, K., Jones, P. R., Smith, L., Nardini, M., & Spence, C. (2015). Hearing where the eyes see: Children use an irrelevant visual cue when localizing sounds. *Child Development, 86*(5), 1449–1457. https://doi.org/10.1111/cdev.12397

Petrini, K., Remark, A., Smith, L., & Nardini, M. (2014). When vision is not an option: Children's integration of auditory and haptic information is suboptimal. *Developmental Science, 17*(3), 376–387. https://doi.org/10.1111/desc.12127

Peykarjou, S., Hoehl, S., & Pauen, S. (2024). The development of visual categorization based on high-level cues. *Child Development*, *95*(2), e122-e138. https://doi.org/10.1111/cdev.14015

Pietrini, P., Furey, M. L., Ricciardi, E., et al. (2004). Beyond sensory images: Object-based representation in the human ventral pathway. *PNAS, 101*(15), 5658–5663. https://doi.org/10.1073/pnas.0400707101

Pike, M. G., Holmstrom, G., De Vries, L. S., Pennock, J. M., Drew, K. J., Sonksen, P. M., & Dubowitz, L. M. S. (1994). Patterns Of Visual Impairment Associated With Lesions Of The Preterm Infant Bran. *Developmental Medicine & Child Neurology*, *36*(10), 849-862. https://doi.org/10.1111/j.1469-8749.1994.tb11776.x

Plebanek, D. J., & Sloutsky, V. M. (2017). Costs of selective attention: When children notice what adults miss. *Psychological Science*, *28*(6), 723-732. https://doi.org/10.1177/0956797617693005

References

Porcu, E., Keitel, C., & Müller, M. M. (2014). Visual, auditory and tactile stimuli compete for early sensory processing capacities within but not between senses. *Neuroimage*, *97*, 224-235. https://doi.org/10.1016/j.neuroimage.2014.04.024

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*(3), 353–363. https://doi.org/10.1037/h0025953

Poulin-Dubois, D., Crivello, C., & Wright, K. (2015). Biological motion primes the animate/inanimate distinction in infancy. *PloS one, 10*(2), e0116910. https://doi.org/10.1371/journal.pone.0116910

Poulin-Dubois, D., Lepage, A., & Ferland, D. (1996). Infants' concept of animacy. *Cognitive Development*, *11*(1), 19-36. https://doi.org/10.1016/S0885-2014(96)90026-X

Pujol, R., & Lavigne-Rebillard, M. (2003). Development and plasticity of the human auditory system. *A Textbook of Audiological Medicine: Clinical Aspects of Hearing and Balance, Dunitz, London*, 147-156.

Pulverman, R., Golinkoff, R. M., Hirsh-Pasek, K., & Buresh, J. S. (2008). Infants discriminate manners and paths in non-linguistic dynamic events. *Cognition*, *108*(3), 825-830. https://doi.org/10.1016/j.cognition.2008.04.009

Purpura, G., Cioni, G., & Tinelli, F. (2018). Development of visuo-haptic transfer for object recognition in typical preschool and school-aged children. *Child Neuropsychology*, *24*(5), 657-670. https://doi.org/10.1080/09297049.2017.1316974

References

Quinn, P. C. (1987). The categorical representation of visual pattern information by young infants. *Cognition, 27*(2), 145–179. https://doi.org/10.1016/0010-0277(87)90017-5

Quinn, P. C. (1994). The categorization of above and below spatial relations by young infants. *Child Development, 65*(1), 58–69. https://doi.org/10.1111/j.1467-8624.1994.tb00734.x

Quinn, P. C. (1999). Development of Recognition and Categorization of Objects and. *Child psychology: A handbook of contemporary issues*, 85.

Quinn, P. C. (2002). Category representation in young infants. *Current Directions in Psychological Science, 11*(2), 66–70. https://doi.org/10.1111/1467-8721.00170

Quinn, P. C., & Eimas, P. D. (1986). On categorization in early infancy. *Merrill-Palmer Quarterly, 32*(4), 331–363. https://doi.org/10.1016/0010-0277(94)90022-1

Quinn, P. C., & Eimas, P. D. (1996). Perceptual cues that permit categorical differentiation of animal species by infants. *Journal of Experimental Child Psychology*, *63*(1), 189-211. https://doi.org/10.1006/jecp.1996.0047

Quinn, P. C., & Johnson, M. H. (2000). Global-before-basic object categorization in connectionist networks and 2-month-old infants. *Infancy*, *1*(1), 31-46.

Quinn, P. C., Doran, M. M., Reiss, J. E., & Hoffman, J. E. (2010). Neural markers of subordinate-level categorization in 6-to 7-month-old infants. *Developmental Science*, *13*(3), 499-507. https://doi.org/10.1111/j.1467-7687.2009.00903.x

Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception*, *22*(4), 463-475. https://doi.org/10.1068/p220463

References

Quinn, P. C., Eimas, P. D., & Tarr, M. J. (2001). Perceptual categorization of cat and dog silhouettes by 3-to 4-month-old infants. *Journal of Experimental Child Psychology*, *79*(1), 78-94. https://doi.org/10.1006/jecp.2000.2609

Quinn, P. C., Slater, A. M., Brown, E., & Hayes, R. A. (2001). Developmental change in form categorization in early infancy. *British Journal of Developmental Psychology*, *19*(2), 207-218. https://doi.org/10.1348/026151001166038

Rabi, R., & Minda, J. P. (2014). Rule-based category learning in children: The role of age and executive functioning. *PloS one*, *9*(1), e85316. https://doi.org/10.1371/journal.pone.0085316

Rakison, D. H., & Butterworth, G. E. (1998). Infants' use of object parts in early categorization. *Developmental Psychology*, *34*(1), 49. https://doi.org/10.1037/0012-1649.34.1.49

Rakison, D. H., & Oakes, L. M. (Eds.). (2003). *Early category and concept development: Making sense of the blooming, buzzing confusion*. Oxford University Press.

Ralph, M. A. L., Lowe, C., & Rogers, T. T. (2007). Neural basis of category-specific semantic deficits for living things: evidence from semantic dementia, HSVE and a neural network model. *Brain: A Journal of Neurology*, *130*(4). https://doi.org/10.1093/brain/awm025

Ramscar, M., & Gitcho, N. (2007). Developmental change and the nature of learning in childhood. *Trends in cognitive sciences*, *11*(7), 274-279. https://doi.org/10.1016/j.tics.2007.05.007

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of “what” and “where” in auditory cortex. *PNAS, 97*(22), 11800–11806. https://doi.org/10.1073/pnas.97.22.11800

References

Rauss, K., & Pourtois, G. (2013). What is bottom-up and what is top-down in predictive coding?. *Frontiers in psychology*, *4*, 276. https://doi.org/10.3389/fpsyg.2013.00276

Reetzke, R., Maddox, W. T., & Chandrasekaran, B. (2016). The role of age and executive function in auditory category learning. *Journal of Experimental Child Psychology*, *142*, 48-65. https://doi.org/10.1016/j.jecp.2015.09.018

Rehder, B., & Hoffman, A. B. (2005). Eye movements and selective attention in category learning. *Cognitive Psychology, 51*(1), 1–41. https://doi.org/10.1016/j.cogpsych.2004.11.001

Rentschler, I., Jüttner, M., Osman, E., Müller, A., & Caelli, T. (2004). Development of configural 3D object recognition. *Behavioural Brain Research*, *149*(1), 107-111. https://doi.org/10.1016/S0166-4328(03)00194-3

Riesenhuber, M., Dickinson, S. J., Leonardis, A., Schiele, B., & Tarr, M. J. (2009). Object categorization in man, monkey, and machine: Some answers and some open questions. *Object Categorization: Computer and Human Vision Perspectives*, 216-240. https://doi.org/10.1017/CBO9780511635465.013

Roark, C. L., Lescht, E., Hampton Wray, A., & Chandrasekaran, B. (2023). Auditory and visual category learning in children and adults. *Developmental Psychology*, *59*(5), 963. https://doi.org/10.1037/dev0001525

Roark, C. L., Thakkar, V., Chandrasekaran, B., & Centanni, T. M. (2024). Auditory category learning in children with dyslexia. *Journal of Speech, Language, and Hearing Research*, *67*(3), 974-988. https://doi.org/10.1044/2023_JSLHR-23-00361

Roberts, K., Jentzsch, I., & Otto, T. U. (2024). Semantic congruency modulates the speed-up of multisensory responses. *Scientific Reports*, *14*(1), 567. https://doi.org/10.1038/s41598-023-50674-4

Robinson, C. W., & Sloutsky, V. M. (2004a). Auditory dominance and its change in the course of development. *Child Development*, *75*(5), 1387-1401. https://doi.org/10.1111/j.1467-8624.2004.00747.x

Robinson, C. W., & Sloutsky, V. M. (2004b). The effect of stimulus familiarity on modality dominance. In *Proceedings of the Annual Meeting of the Cognitive Science Society 26*(26). https://escholarship.org/uc/item/7cj6j8w6

Robinson, C. W., & Sloutsky, V. M. (2013). When audition dominates vision: evidence from cross-modal statistical learning. *Experimental Psychology*, *60*(2), 113-121. https://doi.org/10.1027/1618-3169/a000177

Robinson, C. W., & Sloutsky, V. M. (2019). Two mechanisms underlying auditory dominance: Overshadowing and response competition. *Journal of Experimental Child Psychology*, *178*, 317-340.https://doi.org/10.1016/j.jecp.2018.10.001

Röder, B., & Neville, H. J. (2003). Developmental functional plasticity. In M. H. Johnson (Ed.), *Functional brain development and plasticity* (pp. 209–234). Blackwell Publishing.

Röder, B., Rösler, F., & Neville, H. J. (2004). Event-related potentials during auditory language processing in congenitally blind and sighted people. *Neuropsychologia, 42*(7), 1079–1086. https://doi.org/10.1016/S0028-3932(00)00057-9

Röder, B., Teder-Sälejärvi, W., Sterr, A., Rösler, F., Hillyard, S. A., & Neville, H. J. (1999). Improved auditory spatial tuning in blind humans. *Nature, 400*(6740), 162–166. https://doi.org/10.1038/22106

Rohlf, S., Li, L., Bruns, P., & Röder, B. (2020). Multisensory integration develops prior to cross-modal recalibration. *Current Biology, 30*(9), 1726–1732.e7. https://doi.org/10.1016/j.cub.2020.02.048

Rosch, E. (1976). Classification of real-world objects: Origins and representations in cognition. In R. Bruner et al. (Eds.), *Perspectives on learning and memory* (pp. 327–364). Lawrence Erlbaum.

Rosch, E. (1978). Principles of categorization. *Cognition and Categorization/Erlbaum*. https://doi.org/10.1016/b978-1-4832-1446-7.50028-5

Rosch, E. (1999). Reclaiming concepts. *Journal of Consciousness Studies*, *6*(11-12), 61-77.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573-605. https://doi.org/10.1016/0010-0285(75)90024-9

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382-439. https://doi.org/10.1016/0010-0285(76)90013-X

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, *69*, 181-203. https://doi.org/10.1146/annurev-psych-122216-011805

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926-1928. https://doi/10.1126/science.274.5294.1926

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2006). A more rational model of categorization. *Psychological Review*, 113(4), 1222–1246.

References

Sann, C., & Streri, A. (2007). Perception of object shape and texture in human newborns: Evidence from cross-modal transfer tasks. *Developmental Science, 10*(3), 399–410. https://doi.org/10.1111/j.1467-7687.2007.00593.x

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. https://doi.org/10.1037/0096-1523.33.6.1311

Scheller, M., Garcia, S., Gori, M., & Petrini, K. (2019). Active touch facilitates object size perception in children but not adults: An ERP study. *Brain Research, 1714*, 70–81. https://doi.org/10.1016/j.brainres.2019.146381

Scheller, M., Proulx, M. J., De Haan, M., Dahlmann-Noor, A., & Petrini, K. (2021). Late- but not early-onset blindness impairs the development of audio-haptic multisensory integration. *Developmental Science*, *24*(1), e13001. https://doi.org/10.1111/desc.13001

Schendan, H. E., & Ganis, G. (2015). Top-down modulation of visual processing and knowledge after 250 ms supports object constancy of category decisions. *Frontiers in Psychology*, *6*, 1289. https://doi.org/10.3389/fpsyg.2015.01289

Scherf, K. S., Behrmann, M., Kimchi, R., & Luna, B. (2009). Emergence of global shape processing continues through adolescence. *Child Development, 80*(1), 162–177. https://doi.org/10.1111/j.1467-8624.2008.01252.x

Schmidt, F., Paulun, V. C., van Assen, J. J. R., & Fleming, R. W. (2017). Inferring the stiffness of unfamiliar objects from optical, shape, and motion cues. *Journal of Vision*, *17*(3), 18-18. https://doi.org/10.1167/17.3.18

References

Schneider, T. R., Engel, A. K., & Debener, S. (2008). Multisensory identification of natural objects in a two-way crossmodal priming paradigm. *Experimental Psychology*, *55*(2), 121-132. https://doi.org/10.1027/1618-3169.55.2.121

Schwarzer, G., Küfer, I., & Wilkening, F. (1999). Learning categories by touch: On the development of holistic and analytic processing. *Memory & Cognition*, *27*(5), 868-877. https://doi.org/10.3758/BF03198539

Seitz, A. R., Kim, R., & Shams, L. (2006). Sound facilitates visual learning. *Current Biology, 16*(14), 1422–1427. https://doi.org/10.1016/j.cub.2006.05.048

Sekiyama, K., & Burnham, D. (2008). Impact of language on development of auditory–visual speech perception: A McGurk effect study. *Developmental Science, 11*(2), 306–320. https://doi.org/10.1111/j.1467-7687.2008.00677.x

Serra, M. J., & DeYoung, C. M. (2023). The animacy advantage in memory occurs under self-paced study conditions, but participants' metacognitive beliefs can deter it. *Frontiers in Psychology*, *14*, 1164038. https://doi.org/10.3389/fpsyg.2023.1164038

Setti, A., & Newell, F. N. (2010). The effect of body and part-based motion on the recognition of unfamiliar objects. *Visual Cognition*, *18*(3), 456-480. https://doi.org/10.1080/13506280902830561

Sha, L., Haxby, J. V., Abdi, H., Guntupalli, J. S., Oosterhof, N. N., Halchenko, Y. O., & Connolly, A. C. (2015). The animacy continuum in the human ventral vision pathway. *Journal of Cognitive Neuroscience*, *27*(4), 665-678. https://doi.org/10.1162/jocn_a_00733

Shafiro, V., Sheft, S., Norris, M., Spanos, G., Radasevich, K., Formsma, P., & Gygi, B. (2016). Toward a nonspeech test of auditory cognition: Semantic context effects in environmental sound identification in adults of varying age and hearing

abilities. *PloS one*, *11*(11), e0167030. https://doi.org/10.1371/journal.pone.0167030

Shaikh, D. (2022). Learning multisensory cue integration: A computational model of crossmodal synaptic plasticity enables reliability-based cue weighting by capturing stimulus statistics. *Frontiers in Neural Circuits*, *16*, 921453. https://doi.org/10.3389/fncir.2022.921453

Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Cognitive Sciences, 15*(3), 114–121. https://doi.org/10.1016/j.tins.2010.11.002

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences, 12*(11), 411–417. https://doi.org/10.1016/j.tics.2008.07.006

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*, *75*(13), 1. https://psycnet.apa.org/doi/10.1037/h0093825

Shipley, T. (1964). Auditory flutter–driving of visual flicker. *Science, 145*(3638), 1328–1330. https://doi.org/10.1126/science.145.3638.1328

Shuwairi, S. M., Albert, M. K., & Johnson, S. P. (2007). Discrimination of possible and impossible objects in infancy. *Psychological Science, 18*(4), 303–307. https://doi.org/10.1111/j.1467-9280.2007.01893.x

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking.* Oxford University Press. https://doi.org/10.1093/oso/9780195077872.001.0001

Sifre, R., Olson, L., Gillespie, S., Klin, A., Jones, W., & Shultz, S. (2018). A longitudinal investigation of preferential attention to biological motion in 2- to 24-month-old infants. *Scientific Report*s, 8, 2527. https://doi.org/10.1038/s41598-018-20808-0

Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences*, 105(2), 809-813. https://doi.org/10.1073/pnas.0707021105

Singer, J. J., Karapetian, A., Hebart, M. N., & Cichy, R. M. (2023). The link between visual representations and behavior in human scene perception. *BioRxiv*, 2023-08. Now published in *Imaging Neuroscience* https://doi.org/10.1162/imag_a_00449

Sinnott, J. M., & Aslin, R. N. (1985). Frequency and intensity discrimination in human infants and adults. *The Journal of the Acoustical Society of America*, *78*(6), 1986-1992. https://doi.org/10.1121/1.392655

Siu, C. R., & Murphy, K. M. (2018). The development of human visual cortex and clinical implications. *Eye and Brain*, 25-36. https://doi.org/10.2147/EB.S130893

Slater, A., Mattock, A., Brown, E., & Bremner, J. G. (1991). Form perception at birth: Revisited. *Journal of Experimental Child Psychology*, *51*(3), 395-406. https://doi.org/10.1016/0022-0965(91)90084-6

Slater, A., Morison, V., & Somers, M. (1988). Orientation discrimination and cortical function in the human newborn. *Perception*, *17*(5), 597-602. https://doi.org/10.1068/p170597

Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences*, *7*(6), 246-251. https://doi.org/10.1016/S1364-6613(03)00109-8

Sloutsky, V. M. (2010). From perceptual categories to concepts: What develops? *Cognitive Science, 34*(7), 1244–1286. https://doi.org/10.1111/j.1551-6709.2010.01129.x

References

Sloutsky, V. M. (2016). Selective attention, diffused attention, and the development of categorization. *Cognitive Psychology*, *91*, 24-62. https://doi.org/10.1016/j.cogpsych.2016.09.002

Sloutsky, V. M., & Deng, W. (2017). Categories, concepts, and conceptual development. *Language, Cognition and Neuroscience*, *34*(10), 1284-1297. https://doi.org/10.1080/23273798.2017.1391398

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General, 133*(2), 166–188. https://doi.org/10.1037/0096-3445.133.2.166

Sloutsky, V. M., & Fisher, A. V. (2008). Attentional learning and flexible induction: How mundane mechanisms give rise to smart behaviors. *Child Development*, *79*(3), 639-651. https://doi.org/10.1111/j.1467-8624.2008.01148.x

Sloutsky, V. M., & Fisher, A. V. (2012). Linguistic labels: Conceptual markers or object features? *Journal of Experimental Child Psychology, 111*(1), 65–86. https://doi.org/10.1016/j.jecp.2011.07.007

Sloutsky, V. M., & Napolitano, A. C. (2003). Is a picture worth a thousand words? Preference for auditory modality in young children. *Child Development, 74*(3), 822–833. https://doi.org/10.1111/1467-8624.00570

Sloutsky, V. M., & Robinson, C. W. (2008). The role of words and sounds in visual processing: From overshadowing to attentional tuning. *Cognitive Science*, *32*, 354–377. https://doi.org/10.1080/03640210701863495

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Harvard University Press. https://doi.org/10.4159/harvard.9780674866270

Smith, L. B. (1989). A model of perceptual classification in children and adults. *Psychological Review*, *96*(1), 125. http://dx.doi.org/10.1037/0033-295X.96.1.125

Smith, L. B. (2009). From fragments to geometric shape: Changes in visual object recognition between 18 and 24 months. *Current Directions in Psychological Science*, *18*(5), 290-294. https://doi.org/10.1111/j.1467-8721.2009.01654.x

Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: A dumb attentional mechanism?. *Cognition*, *60*(2), 143-171. https://doi.org/10.1016/0010-0277(96)00709-3

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science, 13*(1), 13–19. https://doi.org/10.1111/1467-9280.00403

Smith, N. A., Folland, N. A., & Trainor, L. J. (2017). 4-month-olds perceive a mistuned harmonic as a separate auditory and visual object. *Cognition, 161*, 24–29. https://doi.org/10.1016/j.cognition.2017.01.016

Sokol, S. (1978). Measurement of infant visual acuity from pattern reversal evoked potentials. *Vision Research, 18*(1), 33–39. https://doi.org/10.1016/0042-6989(78)90074-3

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science*, *14*(1), 29-56. https://doi.org/10.1016/0364-0213(90)90025-R

Spence, C. (2016). Making sense of touch: a multisensory approach to the perception of objects. In *The Power of Touch* (pp. 45-61). Routledge.

Spence, C. (2020). Simple and complex crossmodal correspondences involving audition. *Acoustical Science and Technology*, *41*(1), 6-12. https://doi.org/10.1250/ast.41.6

References

Spencer, J., O'Brien, J., Johnston, A., & Hill, H. (2006). Infants' discrimination of faces by using biological motion cues. *Perception*, *35*(1), 79-89. https://doi.org/10.1068/p5379

Spriet, C., Abassi, E., Hochmann, J. R., & Papeo, L. (2020). Visual object categorization in infancy. *Journal of Vision*, *20*(11), 1079-1079. https://doi.org/10.1073/pnas.2105866119

Stack, D. M., & Tsonis, M. (1999). Infants' haptic perception of texture in the presence and absence of visual cues. *British Journal of Developmental Psychology*, *17*(1), 97-110. https://doi.org/10.1348/026151099165177

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. MIT Press.

Stein, B. E., Labos, E., & Kruger, L. (1973). Sequence of changes in properties of neurons of superior colliculus of the kitten during maturation. *Journal of Neurophysiology*, *36*(4), 667-679. https://doi.org/10.1152/jn.1973.36.4.667

Stein, B. E., Stanford, T. R., & Rowland, B. A. (2014). Development of multisensory integration from the perspective of the individual neuron. *Nature Reviews Neuroscience, 15*(8), 520–535. https://doi.org/10.1038/nrn3742

Stein, B. E., Stanford, T. R., & Rowland, B. A. (2020). Multisensory integration and the society for neuroscience: Then and now. *Journal of Neuroscience*, *40*(1), 3-11. https://doi.org/10.1523/JNEUROSCI.0737-19.2019

Stein, B. E., Wallace, M. T., & Stanford, T. R. (1999). Development of multisensory integration: transforming sensory input into motor output. *Mental Retardation and Developmental Disabilities Research Reviews*, *5*(1), 72-85. https://doi.org/10.1002/(SICI)1098-2779(1999)5:1<72::AID-MRDD8>3.0.CO;2-U

References

Stevenson, R. A., Segers, M., Ferber, S., Barense, M. D., & Wallace, M. T. (2014). The impact of multisensory integration deficits on speech perception in children with autism spectrum disorders. *Frontiers in Psychology, 5*, 379. https://doi.org/10.3389/fpsyg.2014.00379

Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory temporal integration in autism spectrum disorders. *Journal of Neuroscience, 38*(43), 9215–9226. https://doi.org/10.1523/JNEUROSCI.3615-13.2014

Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *Journal of Experimental Psychology: Human Perception and Performance, 38*(6), 1517–1529. https://doi.org/10.1037/a0027339

Streri, A. (2005). Touching for knowing: Cognitive psychology of haptic perception. *John Benjamins Publishing.*

Streri, A., & Gentaz, E. (2003). Cross-modal recognition of shape from hand to eyes in human newborns. *Somatosensory & Motor Research*, *20*(1), 13-18. https://doi.org/10.1080/0899022031000083799

Streri, A., & Gentaz, E. (2004). Cross-modal recognition of shape from hand to eyes and handedness in human newborns. Neuropsychologia, 42(10), 1365-1369. https://doi.org/10.1016/j.neuropsychologia.2004.02.012

Streri, A., Coulon, M., & Guellai, B. (2012). The foundations of social cognition: Studies on face/voice integration in newborn infants. *International Journal of Behavioral Development*, *37*(2), 79-83. https://doi.org/10.1177/0165025412465361

Streri, A., Lhote, M., & Dutilleul, S. (2000). Haptic perception in newborns. *Developmental Science*, *3*(3), 319-327. https://doi.org/10.1111/1467-7687.00126

Striano, T., & Bushnell, E. W. (2005). Haptic perception of texture in 3-month-old infants. *Infant Behavior and Development, 28*(1), 29–38. https://doi.org/10.1016/j.infbeh.2005.05.008

Sun, M., & Savarese, S. (2021). Model-Based Object Recognition: Traditional Approach. In *Computer Vision: A Reference Guide* (pp. 807-812). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-03243-2_334-1

Talkington, W. J., Taglialatela, J. P., & Lewis, J. W. (2013). Using naturalistic utterances to investigate vocal communication processing and development in human and non-human primates. *Hearing Research*, *305*, 74-85. https://doi.org/10.1016/j.heares.2013.08.009

Talsma, D. (2015). Predictive coding and multisensory integration: an attentional account of the multisensory mind. *Frontiers in Integrative Neuroscience*, *9*, 19. https://doi.org/10.3389/fnint.2015.00019

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in cognitive sciences*, *14*(9), 400-410. https://doi.org/10.1016/j.tics.2010.06.008

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23*(3), 457–482. https://doi.org/10.1016/0010-0285(91)90016-H

Taniguchi, K., Tanabe-Ishibashi, A., & Itakura, S. (2020). The categorization of objects with uniform texture at superordinate and living/non-living levels in infants: An

exploratory study. *Frontiers in Psychology*, 11, 2009. https://doi.org/10.3389/fpsyg.2020.02009

Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology: Human Perception and Performance, 21*(6), 1494–1505. https://doi.org/10.1037/0096-1523.21.6.1494

Tassinari, H., Hudson, T. E., & Landy, M. S. (2006). Combining priors and noisy visual cues in a rapid pointing task. *Journal of Neuroscience, 26*(40), 10154–10163.* https://doi.org/10.1523/JNEUROSCI.2779-06.2006

Taylor, G., Hipp, D., Moser, A., Dickerson, K., & Gerhardstein, P. (2014). The development of contour processing: evidence from physiology and psychophysics. *Frontiers in Psychology*, *5*, 719. https://doi.org/10.3389/fpsyg.2014.00719

Taylor, K. I., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2006). Binding crossmodal object features in perirhinal cortex. *Proceedings of the National Academy of Sciences*, *103*(21), 8239-8244. https://doi.org/10.1073/pnas.0509704103

Teki, S., Chait, M., Kumar, S., & Griffiths, T. D. (2013). Segregation of complex acoustic scenes based on temporal coherence. *eLife, 2*, e00699. https://doi.org/10.7554/eLife.00699

Teller, D. Y., Peeples, D. R., & Sekel, M. (1978). Discrimination of chromatic from white light by two-month-old human infants. *Vision Research*, *18*(1), 41-48. https://doi.org/10.1016/0042-6989(78)90075-5

Thelen, A., Talsma, D., & Murray, M. M. (2015). Single-trial multisensory memories affect later auditory and visual object discrimination. *Cognition*, *138*, 148-160. https://doi.org/10.1016/j.cognition.2015.02.003

References

Thomas, R. L., Nardini, M., & Mareschal, D. (2017). The impact of semantically congruent and incongruent visual information on auditory object recognition across development. *Journal of Experimental Child Psychology*, *162*, 72-88. https://doi.org/10.1016/j.jecp.2017.04.020

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. N*ature*, *381*(6582), 520-522. https://doi.org/10.1038/381520a0

Tortelli, C., Senna, I., Binda, P., & Ernst, M. O. (2023). Development of local-global preference in vision and haptics. *Journal of Vision*, *23*(4), 6-6. https://doi.org/10.1167/jov.23.4.6

Tozawa, J., & Oyama, T. (2006). Effects of motion parallax and perspective cues on perceived size and distance. *Perception*, *35*(8), 1007-1023. https://doi.org/10.1068/p5251

Trainor, L. J., & Trehub, S. E. (1992). A comparison of infants' and adults' sensitivity to western musical structure. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(2), 394. https://doi.org/10.1037//0096-1523.18.2.394

Träuble, B., Pauen, S., & Poulin-Dubois, D. (2014). Speed and direction changes induce the perception of animacy in 7-month-old infants. *Frontiers in Psychology*, *5*, 1141. https://doi.org/10.3389/fpsyg.2014.01141

Trehub, S. E. (1973). Infants' sensitivity to vowel and tonal contrasts. *Developmental Psychology*, *9*(1), 91. https://doi.org/10.1037/h0034999

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development*, 466-472. https://doi.org/10.2307/1128803

References

Trehub, S. E., Schneider, B. A., & Henderson, J. L. (1995). Gap detection in infants, children, and adults. *The Journal of the Acoustical Society of America*, *98*(5), 2532-2541. https://doi.org/10.1121/1.414396

Troje, N. F. (2013). What is biological motion? Definition, stimuli, and paradigms. *Social perception: Detection and Interpretation of Animacy, Agency, and Intention*, 13-36. https://doi.org/10.7551/mitpress/9780262019279.003.0002

Turati, C., Simion, F., & Zanon, L. (2003). Newborns' perceptual categorization for closed and open geometric forms. *Infancy*, *4*(3), 309-325. https://doi.org/10.1207/S15327078IN0403_01

Turati, C., Simion, F., Milani, I., & Umilta, C. (2002). Newborns' preference for faces: what is crucial?. *Developmental Psychology*, *38*(6), 875. https://doi.org/10.1037/0012-1649.38.6.875

Ujitoko, Y., & Kawabe, T. (2022). Perceptual judgments for the softness of materials under indentation. *Scientific Reports*, *12*(1), 1761. https://doi.org/10.1038/s41598-022-05864-x

Ujitoko, Y., Kaneko, S., Yokosaka, T., & Kawabe, T. (2023). Falling and heaviness: Heaviness judgment for a visual object which users lift up is influenced by the presentation of the object's falling or staying still. *Frontiers in Psychology*, *14*, 1042188. https://doi.org/10.3389/fpsyg.2023.1042188

Upshaw, M. B., & Sommerville, J. A. (2015). Twelve-month-old infants anticipatorily plan their actions according to expected object weight in a novel motor context. *Frontiers in Public Health*, *3*, 32. https://doi.org/10.3389/fpubh.2015.00032

References

Ursino, M., Cuppini, C., & Magosso, E. (2014). Neurocomputational approaches to modelling multisensory integration in the brain: A review. *Neural Networks, 85*, 17–30. https://doi.org/10.1016/j.neunet.2014.08.003

Valenza, E., Leo, I., Gava, L., & Simion, F. (2006). Perceptual completion in newborn human infants. *Child Development*, *77*(6), 1810-1821. https://doi.org/10.1111/j.1467-8624.2006.00975.x

Vanpaemel, W. (2010). Abstraction and model evaluation in category learning. Behavior Research Methods, 42(2), 421–437. https://doi.org/10.3758/BRM.42.2.421

Vanpaemel, W., & Lee, M. D. (2009). Optimal attention and the GCM. Proceedings of the 31st Annual Conference of the Cognitive Science Society, 2992–2997.

Vanpaemel, W., & Storms, G. (2008). In search of abstraction: The varying abstraction model of categorization. *Psychonomic Bulletin & Review, 15*(4), 732–749. https://doi.org/10.3758/PBR.15.4.732

Vanpaemel, W., & Storms, G. (2010). Abstraction and model evaluation in category learning. *Behavior Research Methods*, *42*(2), 421-437. https://doi.org/10.3758/BRM.42.2.421

Vercillo, T., Burr, D., Sandini, G., & Gori, M. (2015). Children do not recalibrate motor-sensory temporal order after exposure to delayed sensory feedback. *Developmental Science*, *18*(5), 703-712. https://doi.org/10.1111/desc.12247

Vercillo, T., Carrasco, C., & Jiang, F. (2017). Age-related changes in sensorimotor temporal binding. *Frontiers in human neuroscience*, *11*, 500. https://doi.org/10.3389/fnhum.2017.00500

References

Verhaar, E., Medendorp, W. P., Hunnius, S., & Stapel, J. C. (2022). Bayesian causal inference in visuotactile integration in children and adults. *Developmental Science*, *25*(3), e13184. https://doi.org/10.1111/desc.13184

Vida, M. D., & Behrmann, M. (2017). Subcortical facilitation of behavioral responses to threat. *Scientific Reports*, *7*(1), 13087.. https://doi.org/10.1038/s41598-017-13203-8

Viganò, S., Vercillo, T., & Koini, M. (2021). Symbolic categorization of novel multisensory stimuli in the human brain. *NeuroImage, 236*, 118110. https://doi.org/10.1016/j.neuroimage.2021.118016

Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: the privileged status of speech for young infants. *Developmental Science*, *7*(3), 270-276. https://doi.org/10.1111/j.1467-7687.2004.00345.x

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception, & Psychophysics, 72*(4), 871–884. https://doi.org/10.3758/APP.72.4.871

Vukatana, E., Zepeda, M. S., Anderson, N., Curtin, S., & Graham, S. A. (2020). Eleven-month-olds link sound properties with animal categories. *Frontiers in Psychology*, *11*, 559390. https://doi.org/10.3389/fpsyg.2020.559390

Wallace, M. T., & Stein, B. E. (1997). Development of multisensory neurons and multisensory integration in cat superior colliculus. *Journal of Neuroscience, 17*(7), 2429–2444. https://doi.org/10.1523/JNEUROSCI.17-07-02429.1997

Wallace, M. T., & Stein, B. E. (2007). Early experience determines how the senses will interact. *Journal of Neurophysiology*, *97*(1), 921-926. https://doi.org/10.1152/jn.00497.2006

References

Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia*, *64*, 105-123. https://doi.org/10.1016/j.neuropsychologia.2014.08.005

Wallace, M. T., Perrault, T. J., Hairston, W. D., & Stein, B. E. (2004). Visual experience is necessary for the development of multisensory integration. *Journal of Neuroscience, 24*(43), 9580–9584. https://doi.org/10.1523/JNEUROSCI.2535-04.2004

Wallace, M. T., Woynaroski, T. G., & Stevenson, R. A. (2020). Multisensory integration as a window into orderly and disrupted cognition and communication. *Annual Review of Psychology*, *71*(1), 193-219. https://doi.org/10.1146/annurev-psych-010419-051112

Warren, D. H., Welch, R. B., & McCarthy, T. J. (1981). The role of visual–auditory "compellingness" in the ventriloquism effect: Implications for transitivity among the spatial senses. *Perception & Psychophysics, 30*(6), 557–564. https://doi.org/10.3758/BF03202010

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain*, *107*(3), 829-853. https://doi.org/10.1093/brain/107.3.829

Watson, E. A., Ewing, L., & Malcolm, G. L. (2025). When children get the gist: The development of rapid scene categorisation. *Vision Research*, *233*, 108620. https://doi.org/10.1016/j.visres.2025.108620

Wattam-Bell, J., Chiu, M., & Kulke, L. (2012). Developmental reorganisation of visual motion pathways. *iPerception*, *3*(4), 230. https://doi.org/10.1068/id230

Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences, 13*(6), 258–263. https://doi.org/10.1016/j.tics.2009.03.006

Wei, H., Dong, Z., & Wang, L. (2018). V4 shape features for contour representation and object detection. *Neural Networks*, *97*, 46-61. https://doi.org/10.1016/j.neunet.2017.09.010

Weichart, E. R., Unger, L., King, N., Sloutsky, V. M., & Turner, B. M. (2024). "The eyes are the window to the representation": Linking gaze to memory precision and decision weights in object discrimination tasks. *Psychological Review*, *131*(4), 1045. https://doi.org/10.1037/rev0000475

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88*(3), 638–667. https://doi.org/10.1037/0033-2909.88.3.638

Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, 349-355. https://doi.org/10.2307/1129249

Werner, L. A. (2019). Development of auditory behavior (Chapter 3). In G. M. Werner & R. R. Fay (Eds.), *The sense of hearing in children* (pp. 49–84). Springer. https://doi.org/10.1007/978-1-4614-1421-6

Werner, L. A., & Boike, K. (2001). Infants' sensitivity to broadband noise. *The Journal of the Acoustical Society of America*, *109*(5), 2103-2111. https://doi.org/10.1121/1.1365112

Werner, L. A., Folsom, R. C., & Mancl, L. R. (1994). The relationship between auditory brainstem response latencies and behavioral thresholds in normal hearing infants

and adults. *Hearing Research*, *77*(1-2), 88-98. https://doi.org/10.1016/0378-5955(94)90256-9

Werner, L. A., Marean, G. C., Halpin, C. F., Spetner, N. B., & Gillenwater, J. M. (1992). Infant auditory temporal acuity: Gap detection. *Child Development*, *63*(2), 260-272. https://doi.org/10.1111/j.1467-8624.1992.tb01625.x

Wightman, F., Allen, P., Dolan, T., Kistler, D., & Jamieson, D. (1989). Temporal resolution in children. *Child Development*, 611-624. https://doi.org/10.2307/1130727

Wilcox T. (1999). Object individuation: infants' use of shape, size, pattern, and color. *Cognition*. *72*(2), 125–166. https://doi.org/10.1016/S0010-0277(99)00035-9

Wilcox, T., & Biondi, M. (2016). Functional activation in the ventral object processing pathway during the first year. *Frontiers in Systems Neuroscience*, *9*, 180. https://doi.org/10.3389/fnsys.2015.00180

Wilcox, T., Stubbs, J., Hirshkowitz, A., & Boas, D. A. (2012). Functional activation of the infant cortex during object processing. *NeuroImage*, *62*(3), 1833-1840. https://doi.org/10.1016/j.neuroimage.2012.05.039

Wille, C., & Ebersbach, M. (2016). Semantic congruency and the (reversed) Colavita effect in children and adults. *Journal of Experimental Child Psychology*, *141*, 23-33. https://doi.org/10.1016/j.jecp.2015.07.015

Williams, J. R., Markov, Y. A., Tiurina, N. A., & Störmer, V. S. (2022). What you see is what you hear: Sounds alter the contents of visual perception. *Psychological Science*, *33*(12), 2109-2122. https://doi.org/10.1177/09567976221121348

Winkler, I., Kushnerenko, E., Horváth, J., Čeponienė, R., Fellman, V., Huotilainen, M., ... & Sussman, E. (2003). Newborn infants can organize the auditory

world. *Proceedings of the National Academy of Sciences*, *100*(20), 11812-11815. https://doi.org/10.1073/pnas.2031891100

Withagen, A., Kappers, A. M., Vervloed, M. P., Knoors, H., & Verhoeven, L. (2012). Haptic object matching by blind and sighted adults and children. *Acta Psychologica*, *139*(2), 261-271. https://doi.org/10.1016/j.actpsy.2011.11.012

Woods, A. T., & Newell, F. N. (2004). Visual, haptic and cross-modal recognition of objects and scenes. Journal of Physiology-Paris, 98(1-3), 147-159. https://doi.org/10.1016/j.jphysparis.2004.03.006

Xie, S., Hoehl, S., Moeskops, M., Kayhan, E., Kliesch, C., Turtleton, B., ... & Cichy, R. M. (2022). Visual category representations in the infant brain. *Current Biology*, *32*(24), 5422-5432. https://doi.org/10.1016/j.cub.2022.11.016

Yakovlev, P. I., & Lecours, A. R. (1967). The myelogenetic cycles of regional maturation of the brain. In A. Minkowski (Ed.), *Regional development of the brain in early life* (pp. 3–70). Blackwell.

Yan, X., Tung, S. S., Fascendini, B., Chen, Y. D., Norcia, A. M., & Grill-Spector, K. (2024). The emergence of visual category representations in infants' brains. *eLife*, *13*, RP100260. https://doi.org/10.7554/eLife.100260.3

Yao, L., Fu, Q., Liu, C. H., Wang, J., & Yi, Z. (2025). Distinguishing the roles of edge, color, and other surface information in basic and superordinate scene representation. *NeuroImage*, *310*, 121100. https://doi.org/10.1016/j.neuroimage.2025.121100

Yi, H. G., Maddox, W. T., Mumford, J. A., & Chandrasekaran, B. (2016). The Role of Corticostriatal Systems in Speech Category Learning. *Cerebral Cortex (New York, NY: 1991)*, *26*(4), 1409-1420. https://doi.org/10.1093/cercor/bhu236

References

Yildirim, I., & Jacobs, R. A. (2013). Transfer of object category knowledge across visual and haptic modalities: Experimental and computational studies. *Cognition, 126*(2), 135–148. https://doi.org/10.1016/j.cognition.2012.08.005

Younger, B. A. (1985). The segregation of items into categories by ten-month-old infants. *Child Development, 56*(6), 1574–1583. https://doi.org/10.2307/1130476

Younger, B. A., & Cohen, L. B. (1986). Developmental change in infants' perception of correlations among attributes. *Child Development, 57*(4), 803–815. https://doi.org/10.2307/1130356

Younger, B. A., & Fearing, D. D. (2000). A global-to-basic trend in early categorization: Evidence from a dual-category habituation task. *Infancy*, *1*(1), 47-58. https://doi.org/10.1207/S15327078IN0101_05

Younger, B., & Gotlieb, S. (1988). Development of categorization skills: Changes in the nature or structure of infant form categories?. *Developmental Psychology*, *24*(5), 611. https://doi.org/10.1037/0012-1649.24.5.611

Yuodelis, C., & Hendrickson, A. (1986). A qualitative and quantitative analysis of the human fovea during development. *Vision Research*, *26*(6), 847-855. https://doi.org/10.1016/0042-6989(86)90143-4

Zanolie, K., & Crone, E. A. (2018). Development of cognitive control across childhood and adolescence. *The Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience,*, *3*, 159-182. https://doi.org/10.1002/9781119170174.epcn405

Zepeda, M. S., & Graham, S. A. (2019). Does category-training facilitate 11-month-olds' acquisition of unfamiliar category-property associations?. *Infant Behavior and Development*, *57*, 101380. https://doi.org/10.1016/j.infbeh.2019.101380

References

Zhang, J., & Rowe, J. B. (2014). Dissociable mechanisms of speed-accuracy tradeoff during visual perceptual learning are revealed by a hierarchical drift-diffusion model. *Frontiers in Neuroscience*, *8*, 69.https://doi.org/10.3389/fnins.2014.00069

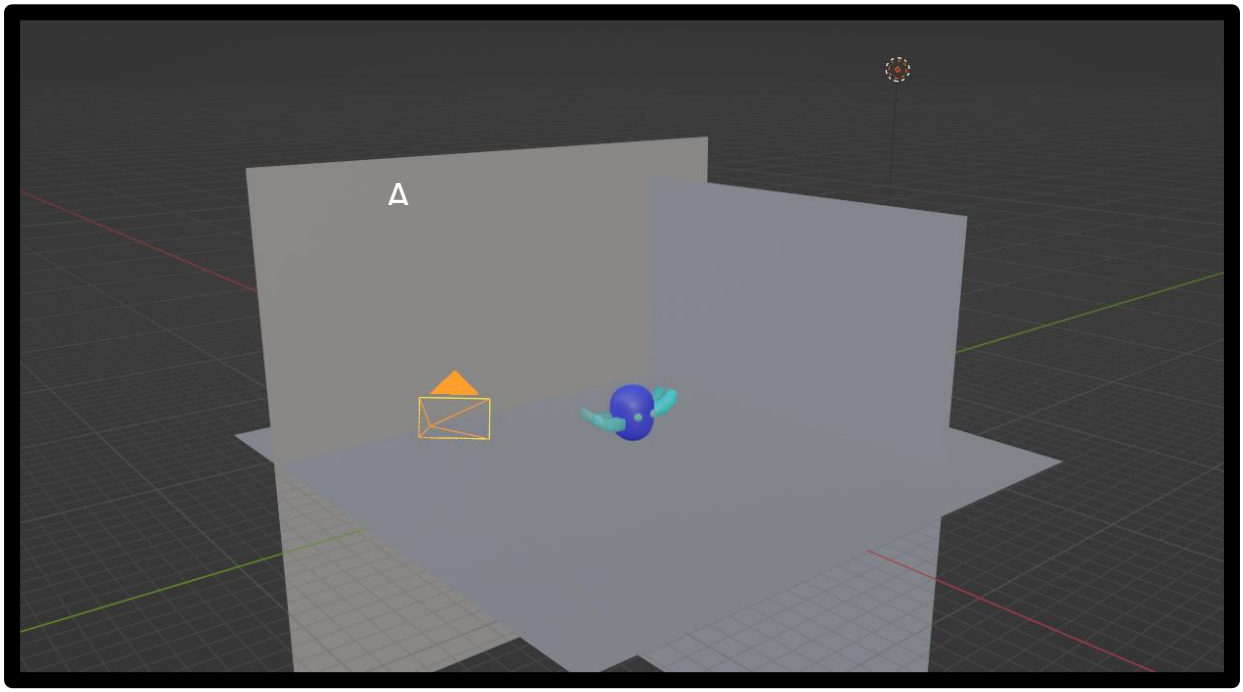
Zhou, H., Schafer, R. J., & Desimone, R. (2016). Pulvinar–cortex interactions in vision and attention. *Neuron, 89*(1), 209–220. https://doi.org/10.1016/j.neuron.2015.11.034

Supplemental Materials

Supplemental 1: Properties of the 3D rendering space for animations.

Figure S1

Object animation rendering space for visual stimuli



Note. An example of the virtual space within each of the target objects were rendered.

The scene depicts the position of the camera (A) and light source (B).

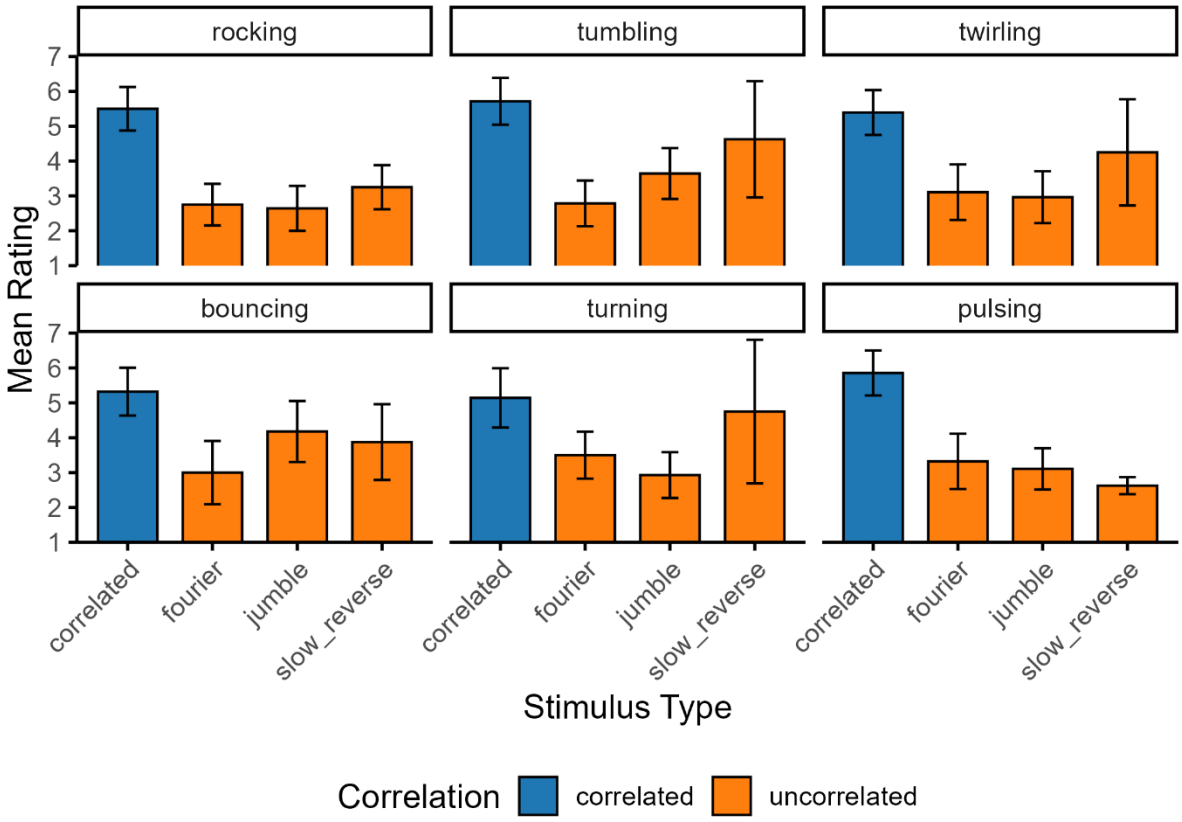
Supplemental 2: Audio-visual correlation validation.

The pilot was conducted online, and participants were recruited using Prolific (*N* = 14). In this study we recruited 14 naive participants with normal or corrected to normal vision and normal hearing who were presented with each visual stimulus -the object movement paired to four different auditory stimuli. These participants were then asked to rate on a 7-point Likert scale whether the video and sound appeared to be correlated (7) or uncorrelated (1). The auditory stimuli were our correlated auditory stimuli, and the Fourier transformed auditory stimulus as well as two other uncorrelated formats, in one we randomly jumbled the original correlated auditory stimulus in 150ms segments, and we also reversed the original audio and altered the temporal information, so it was .35 times slower than the original. We found that the Fourier transform audio was most consistently rated as less correlated than other uncorrelated conditions (see Figure SE2. for details). Participants rated our correlated stimulus (M = 5.49, SD = 1.00) as significantly more correlated to the visual movement than our Fourier auditory stimulus (M = 3.08, SD = 1.12). A paired-samples t-test confirmed this difference was statistically significant [t(13) = 5.86, $p < 0.001$, Cohen's d = 2.27, 95 CI [0.76, 3.78]].

Supplemental Materials

Figure S2

Ratings of Audiovisual correlation across both correlated and uncorrelated stimuli

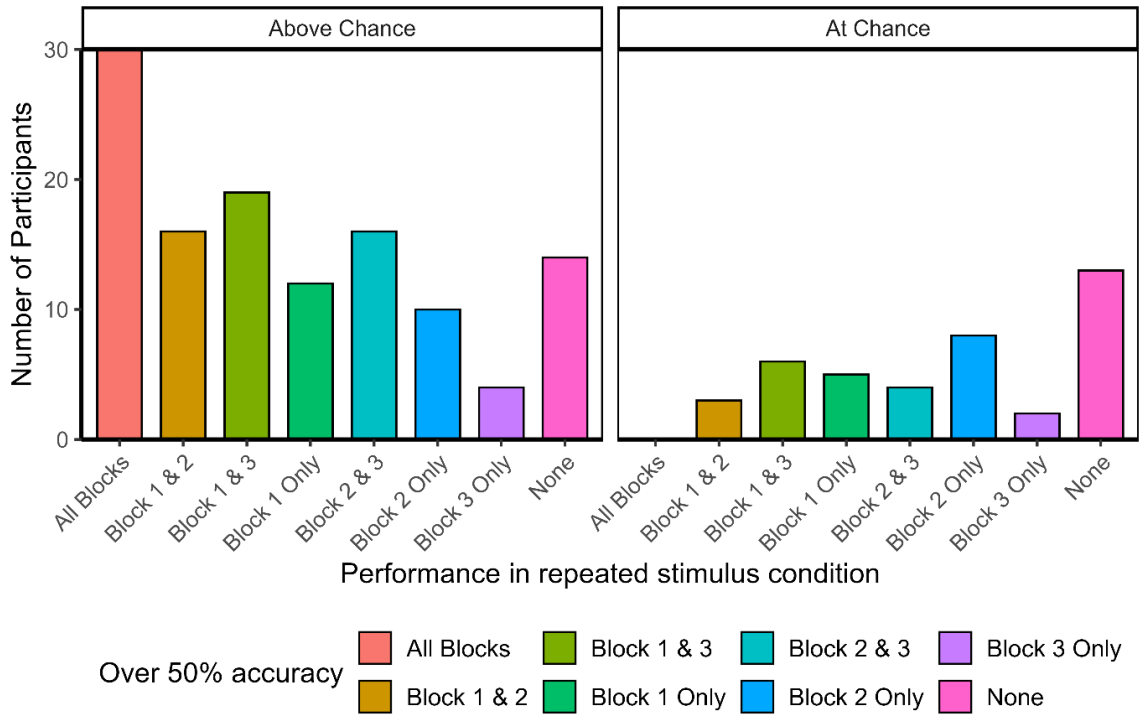


Note. Participant mean ratings on a 7 point Likert scale of whether each auditory stimulus appeared correlated (7) or uncorrelated (1) to the object movement.

Supplemental 3: Figure describing rationale of exclusions for analysis of categorisation test performance.

Figure S3

An assessment of participant overall and per-block accuracy during categorisation test



Note. Panels show the count of participants who scored above (left) and below or at (right) chance during our categorisation test. Note, none refers here to participants who did not score above 50% accuracy in any of the test blocks. Bars represent performance during each repetition (block 1, 2 etc.) of the moving object stimuli (with correlated sound) during the Learning session.

Supplemental 4: Categorisation Accuracy and Reaction Times for the developmental sample (5-13 years) across all movement and sound conditions.

Table S1

Mean Proportion Accuracy and Standard Deviations (sd) averaged across age groups

Movement	Sound	m	sd	N
static	correlated	0.585	0.251	116
static	uncorrelated	0.620	0.216	116
familiar	correlated	0.657	0.199	116
familiar	uncorrelated	0.606	0.232	116
unfamiliar	correlated	0.595	0.248	116
unfamiliar	uncorrelated	0.624	0.218	116

Table S2

Mean log Reaction Times and Standard Deviations (sd) averaged across developmental age groups.

Movement	Sound	m	sd
static	correlated	0.20	1.64
static	uncorrelated	0.16	1.63
familiar	correlated	0.16	1.62
familiar	uncorrelated	0.09	1.70
unfamiliar	correlated	0.20	1.63
unfamiliar	uncorrelated	0.31	1.59

Table S3

Details of the mean proportion accuracy (and standard deviation) for each movement pairing by sound by age group condition. Note the 3-way interaction was not significant.

Age group	Movement Pairings	Sound Condition	Accuracy (*M*)	Accuracy (*SD*)
Child	Learned	Correlated	0.657	0.20
Adult	Familiar	Correlated	0.870	0.19
Child	Familiar	Uncorrelated	0.61	0.23
Adult	Familiar	Uncorrelated	0.82	0.27
Child	Novel	Correlated	0.60	0.25
Adult	Unfamiliar	Correlated	0.80	0.27
Child	Unfamiliar	Uncorrelated	0.62	0.22
Adult	Unfamiliar	Uncorrelated	0.77	0.29
Child	Static	Learned	0.59	0.25
Adult	Static	Learned	0.79	0.29
Child	Static	Uncorrelated	0.62	0.22
Adult	Static	Uncorrelated	0.82	0.24

Supplemental 5: Categorisation Test performance: categorisation reaction times (logRT) across all movement and sound pairings compared between children and adults.

A linear mixed-effects model (LMM) was fit to log-transformed reaction times (log RT) to examine effects of movement pairing (static, learned, novel) and sound type (correlated, uncorrelated), with age group (children aged 5–13 years, adults) as a between-subjects factor. Fixed effects included movement pairings, sounds, age group, and all interactions; a random intercept for participant accounted for repeated measures. Models were estimated using restricted maximum likelihood (REML), and *t*-tests for fixed effects used Satterthwaite's degrees-of-freedom approximation. The scaled residuals were centered (Median = 0.006), with interquartile range from −0.432 to 0.444 and extremes from −3.57 to 3.41, indicating a small number of outlying trials but generally acceptable dispersion. The participant-level random intercept variance was 1.9445 (SD = 1.3945) on the log-RT scale, and the residual variance was 0.1464 (SD = 0.3827), reflecting substantial between-person differences in baseline RT alongside within-person residual variance. The model explained a very high proportion of total variance (conditional R^2 = .938), with fixed effects alone accounting for marginal R^2 = .120. This indicates that the fixed predictors explained ~12% of variance in log RT, while random participant effects captured the remaining variance.

Type III Wald χ^2 tests of the fixed effects on our dependent variable of interest (log RT) indicated a significant main effect of age group [$\chi^2(1) = 19.42$, $p < .001$ $\eta^2p = .12$, 95% CI [0.05, 1.00]], with adults responding faster than children [(estimated difference = −1.277, SE = 0.290); $t(166.76) = -4.41$, $p < .001$], with mean logRT differences across all within subject factor combinations (children M = 0.19, SD = 1.57; adult M = -1.09, SD = 0.33). The main effects of movement pairing [$\chi^2(2) = 0.96$, $p = .618$, $\eta^2p = .005$] and sound [$\chi^2(1) = 1.21$, $p = .272$, $\eta^2p < .001$] were not significant.

None of the pairwise interactions between movement pairings and sound [$\chi^2(2) = 5.38$, $p = .068$], age group and movement pairings [$\chi^2(2) = 0.04$, $p = .980$], age group and sound [$\chi^2(1) = 0.50$, $p = .480$] reached significance. The three-way interaction of our fixed effects was also not significant [$\chi^2(2) = 1.43$, $p = .488$].

Supplemental 6: Stimulus Details

A full overview of the category diagnostic and non-diagnostic features and their different versions are shown below in table S1. Note the distractor feature variants were counterbalanced across the 54 stimuli and ensured an equal number of each variant per category. The diagnostic cues were also counterbalanced across the 12 stimuli used during the learning task, and all test stimuli (i.e. the 6 stimuli from the learning task and the 24 object with *within* category feature change and the 18 objects with *across* category feature change).

Table S4

Different variants of object features both diagnostic and non-diagnostic of category membership

DIAGNOSTIC FEATURES	**VARIANT** (CATEGORY A)	**VARIANT** (CATEGORY B)	**NON-DIAGNOSTIC FEATURES**	**VARIANTS**
VISUAL	Bright	Dark	**VISUAL**	
Leg Colour	Yellow	Purple	Body Colour	red, orange, navy, blue
	Orange	Blue		
HAPTIC	Light	Heavy	**HAPTIC**	
Body Weight	Cushion Filler	Rice	Body Compressibility	Chickpeas, rice, plastic balls, cushion filler
	Plastic Balls	Chickpeas		
VISUOHAPTIC	Soft Malleable	Stiff Spikes	**VISUOHAPTIC**	
Hair Type	Teased wool	Wire Lined Felt	Leg Texture	fabric, mesh lined fabric, felt, ribbon
	Wool	Pipecleaners		
VISUOHAPTIC	Soft	Rough	**VISUOHAPTIC**	
Patches	Felt	Mesh	Body Shape	hemi-ovoid, hemi-sphere, square based pyramid, triangular based pyramid
	Ribbon	Hessian		

Supplemental Materials

Distractor features: Body colour was a visual only distractor feature varying from blue, navy, red and orange. In addition, the haptic only distractor feature was body compressibility which varied along two classifications, hard and soft with the cushion filler cotton and rice classed as compressible, and the chickpeas and plastic balls classed as hard. This haptic-only cue was counterbalanced to the weight of the objects. The distractor visuohaptic feature of stimulus leg texture had four variants (ribbon, felt, fabric, and mesh-lined fabric). Finally, body shape was another visuohaptic distractor stimulus feature and included a triangle-based pyramid, a square-based pyramid, a hemisphere, and a hemi-ovoid.

Supplemental 7: Participant Instructions

All participants were greeted and asked their age in years (in order to assign them to the correct learning condition for counterbalancing purposes. They were then asked to sit at the table opposite to the experimenter where the following participant task instructions were used:

"*In this game, it is your mission to try get these creatures back home. They crash landed on earth, got all mixed up, and they do not get along very well. We need your help to try to figure out which ones are the Neems, and which ones are the Dorps so that we can put them back into their correct spaceships (gesture to indicate the relevant spaceship/box). It is your mission to solve the puzzle of what makes a Neem different to a Dorp. In the beginning, it might be tricky, but this is all about learning, so I just want you to make your best guess. How does that sound - do you have any questions? Are you ready to start your mission?"*

Supplemental 8: Category Learning Task: Learning Outcome

Supplemental analysis were conducted on the effect of learning modality and age group on the probability of passing the learning phase of the experiment.

A binary logistic regression was conducted to examine the effects of age group (4-6 years, 7-9 years, 10-13 years) and learning modality (visual, haptic, visuohaptic) on the likelihood of successful category learning. The overall model was statistically significant, $\chi^2(8) = 63.74$, $p < .001$, indicating that age group and learning modality influenced learning outcomes. The model explained 20.1% (Nagelkerke R^2) of the variance in passing rates. A significant main effect of age group was found ($\chi^2(2) = 41.82$, $p < .001$). Post-hoc comparisons revealed that children aged 7-9 years (OR = 2.88, $p = .045$) and 10-13 years (OR = 2.88, $p = .045$) were significantly more likely to pass compared to 4-6-year-olds. A significant main effect of modality was observed ($\chi^2(2) = 7.19$, $p = .028$), though individual condition effects were not statistically significant. There was a significant age group × modality interaction ($\chi^2(4) = 14.73$, $p = .005$), suggesting that the effect of modality depended on age. Specifically, children aged 10-13 years in the visuohaptic condition were significantly more likely to pass (OR = 6.60, $p = .045$) compared to the 4–6-year-old children in the visuohaptic condition. A visualization of the pass and fail rates across participants is presented in S1.

Figure S4

Descriptive plot depicting number of participants who passed (in green) and failed (in red) the Category Learning task across the different learning modalities and age groups

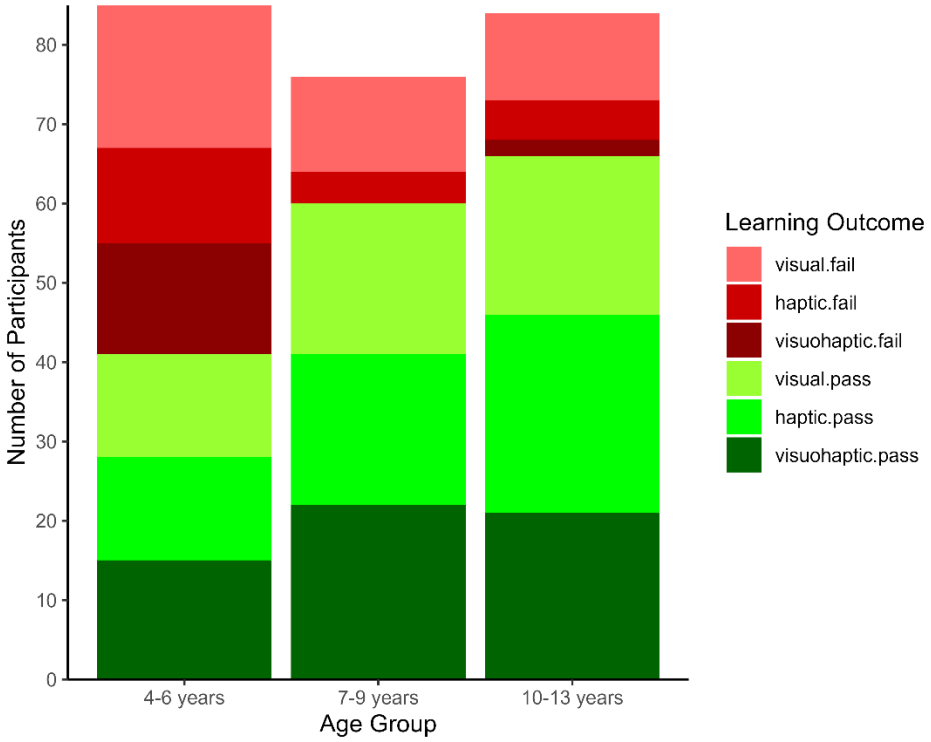


Table S5

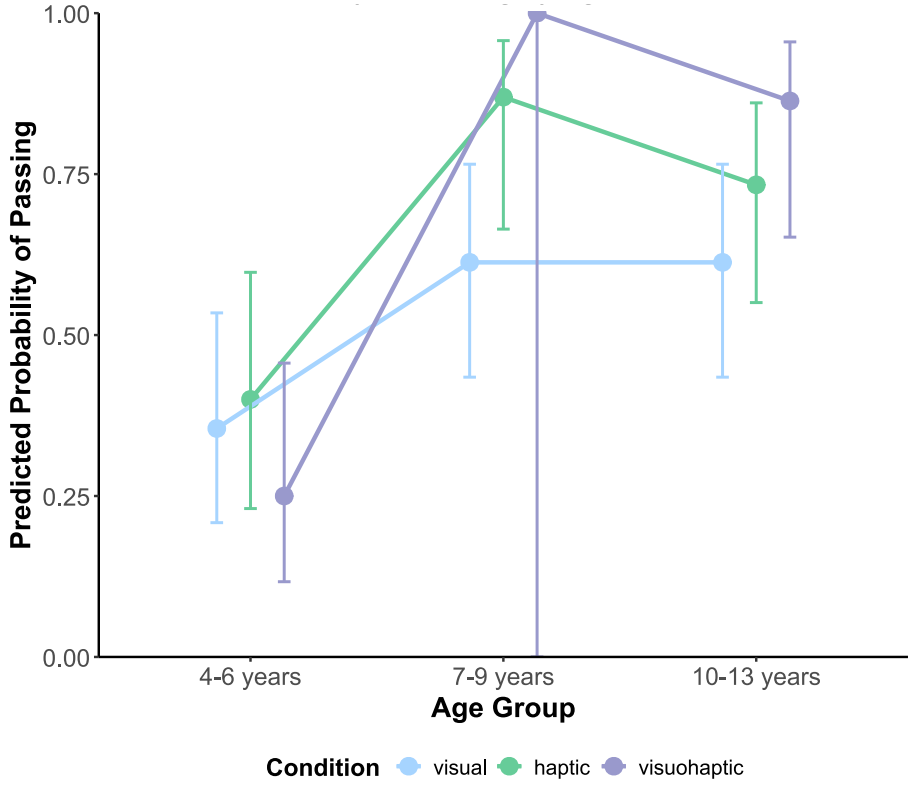
Participant outcomes across learning modality and age groups

Learning modality	Age Group			
	4-6 years (*n*)	7-9 years (*n*)	10+ years (*n*)	Total (*N*)
Visual Pass	13	19	20	52
Visual Fail	18	12	11	41
Visual Total	31	31	31	93
Haptic Pass	13	19	25	57
Haptic Fail	12	4	5	21
Haptic Total	25	23	30	78
Visuohaptic Pass	15	22	21	58
Visuohaptic Fail	14	0	2	16
Visuohaptic Total	29	22	23	74
Total (N)	85	76	84	245

Note. Participants who withdrew: 5 VH and 2 V learners in the 4–6-year-olds.

Figure S5

Predicted probability of passing the learning task across age groups and learning modality.



Category Learning Rate Efficacy

A hierarchical linear mixed effects model was conducted to examine how age group (4–6 years, 7–9 years, 10–13 years) and learning modality (Bimodal, Unimodal Haptic, Unimodal Visual) influenced learning trajectories—cumulative accuracy across trials (normalized trial number). The model was used to capture non-linear learning trajectories while accounting for the effects of individual differences in learning slopes over time.

The model demonstrated good fit to the data (*AIC* = –3467.10, *BIC* = –3334.30, log-likelihood = 1755.50, deviance = –3511.10). The variance of the random intercept (σ^2 = 0.081, *SD* = 0.285) indicated variability in baseline performance across participants. Compared to a random-intercept-only model, the full model provided a significantly better fit, $\chi^2(2) = 893.59$, $p < .001$. The variance of the slope for normalized trial count ($\sigma^2 = 0.096$, *SD* = 0.310) suggested individual differences in learning rates, with a strong negative correlation between intercepts and slopes (r = –0.91), indicating that participants with higher initial accuracy improved less over time. The marginal R^2 was .20, indicating that

fixed effects explained 20% of the variance, while the conditional R^2 was .75, showing that the full model, including random effects, accounted for 75% of variance in cumulative accuracy.

There was a significant main effect of trial number (β = 0.25, SE = 0.09, $t(160)$ = 2.74, p = .007), demonstrating that cumulative accuracy increased over time. No significant main effects of age group were found: 7–9-year-olds (β = 0.02, p = .83) and 10–13-year-olds (β = –0.13, p = .20) did not differ significantly from 4–6-year-olds. Similarly, learning modality did not significantly predict accuracy; neither the Haptic-only (β = –0.09, p = .40) nor the Visual-only (β = –0.10, p = .34) groups differed from the Bimodal condition.

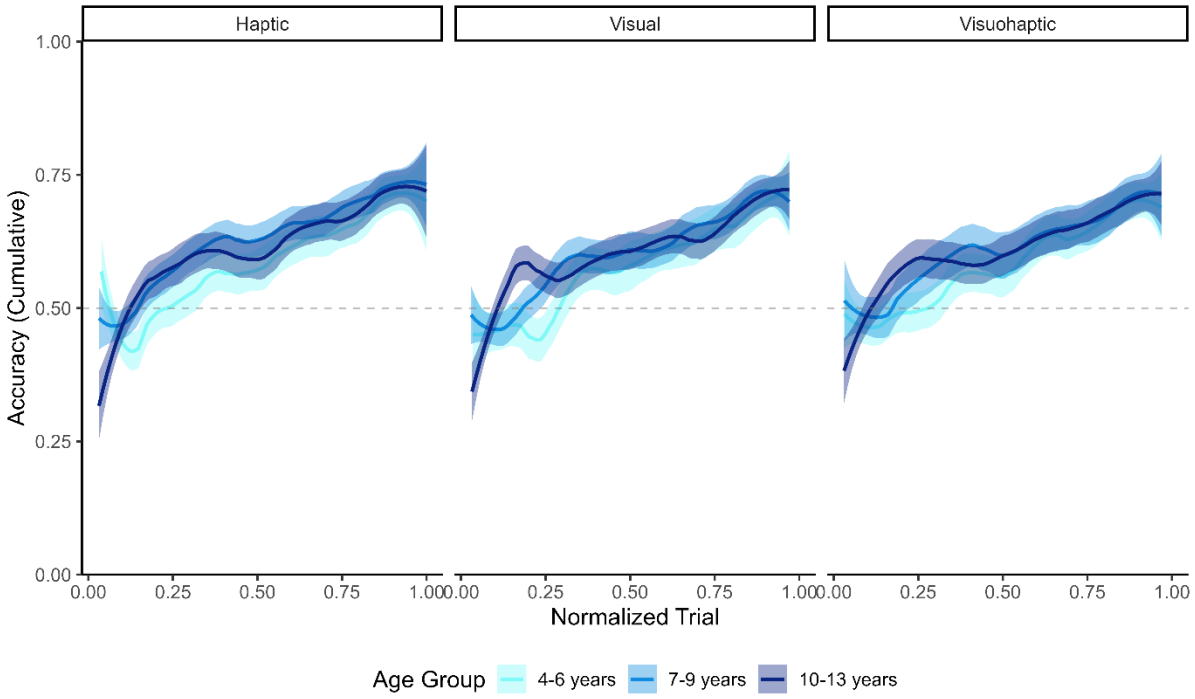
Interactions between normalized trial number and age group revealed a marginal trend for 10–13-year-olds to improve faster over trials (β = 0.21, SE = 0.11, t = 1.89, p = .06). No significant interactions involving learning modality and time were observed. There was no significant three-way interaction observed however, marginal interactions were found comparing learning trajectories of 4–6-year-olds in the Bimodal condition to 10–13-year-olds in the Visual (β = –0.28, SE = 0.16, p = .076) and Bimodal (β = –0.27, SE = 0.16, p = .081) conditions, suggesting weak trends that the oldest children exhibited shallower learning trajectories in these modalities; however, these did not reach significance.

Overall, results indicate that cumulative accuracy improves across trials. There were no significant differences based on age group or learning modality, and learning trajectories were generally stable. Although weak trends suggested that 10–13-year-olds might show modality-specific differences in learning rate, these effects were not statistically significant.

Supplemental Materials

Figure S6

Cumulative accuracy across time (normalised trial) per learning modality and age group



Supplemental 9: Analysis of Categorisation Accuracy for specific cross category feature changes

A 3x3x4 mixed ANOVA was conducted to examine the effects of age group (*4–6 years*, *7–9 years*, *10–13 years*; between-subjects), learning modality (*Haptic*, *Visual*, *Visuohaptic*; between-subjects), and specific cross category feature dimension (*weight (h)*, *leg colour (v)*, *patch (vh)*, *hair (vh)*; within-subjects) on mean proportion accuracy in the categorisation test.

Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of *feature type* [$\chi^2(5) = 0.80$, $p < .001$] as well as for the interactions of age group by feature type, learning modality by feature type, and age group by learning modality by feature type (all *p*s < .001). Therefore, degrees of freedom were corrected where applicable using the Greenhouse–Geisser estimates of sphericity ($\varepsilon = 0.861$).

There was a significant main effect of age group [$F(2, 153) = 14.88$, $p < .001$, *ges* = .045], with post-hoc pairwise comparisons (Tukey-adjusted) indicating that the *4–6 years* group ($M = 0.581$, $SD = 0.19$, 95% CI [0.535, 0.627]) performed significantly worse than both the *7–9 years* group ($M = 0.718$, $SD = 0.13$, 95% CI [0.682, 0.753]; $t(153) = -4.68$, $p < 0.001$), and the *10–13 years* group ($M = 0.728$, $SD = 0.11$, 95% CI [0.695, 0.762]; $t(153) = -5.14$, $p < 0.0001$). The *7–9 years* and *10–13 years* groups did not differ significantly ($t(153) = -0.44$, $p < 0.899$). There was no significant main effect of learning modality [$F(2, 153) = 0.29$, $p = .749$, *ges* < .001].

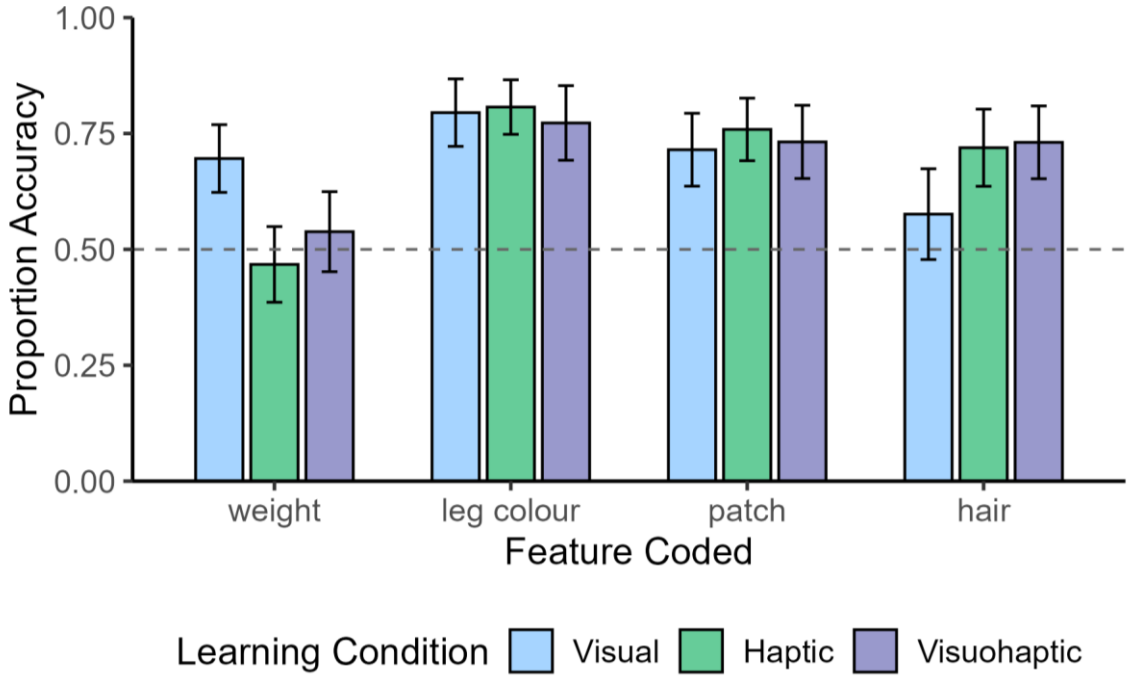
The main effect of feature type was significant [$F(3, 459) = 14.21$, $p < .001$, *ges* = .066], Greenhouse–Geisser corrected. Post-hoc pairwise comparisons showed that *leg colour* ($M = 0.769$, $SD = 0.26$, 95% CI [0.729, 0.809]; $t(153) = -6.51$, $p < 0.001$) and *patch* ($M = 0.707$, $SD = 0.29$, 95% CI [0.666, 0.748]; $t(153) = -4.762$, $p < 0.001$) yielded significantly higher accuracy than *weight* ($M = 0.563$, $SD = 0.31$, 95% CI [0.515, 0.611). *Hair* ($M = 0.664$, $SD = 0.32$, 95% CI [0.613, 0.714]) did not differ significantly from *weight* ($p = .0635$) but was significantly lower than *leg colour* ($t(153) = 3.461$, $p < .01$).

The learning modality by feature type interaction was significant [$F(6, 459) = 4.27$, $p < .001$, *ges* = .041], Greenhouse–Geisser corrected. Inspection of estimated marginal means and post hoc pairwise comparisons indicated that accuracy for a change in the *weight* feature was significantly lower in the Haptic learning condition ($M = 0.468$, $SD = 0.31$, 95% CI [0.387, 0.549]) compared to the Visual learning condition ($M = 0.703$, $SD = 0.27$,

95% CI [0.617, 0.789]; $t(153) = -3{,}93$, $p < 0.01$). For the cross category feature of *hair*, accuracy was high in the Visuohaptic learning ($M = 0.71$, $SD = 0.29$, 95% CI [0.622, 0.795]) and Haptic learning conditions ($M = 0.70$, $SD = 0.32$, 95% CI [0.613, 0.783]) and lowest in the Visual learning condition ($M = 0.58$, $SD = 0.34$, 95% CI [0.493, 0.675]) however, these differences did not reach significance. These findings align with our qualitative assessment of participant category judgements, for more information see S5. None of the age group by learning modality , age group by feature type (both *p*s > .092), nor the three-way ($p = .713$) interactions were significant.

Figure S7

Categorisation accuracy for objects with one cross-category cue compared across learning modality and each type of feature.

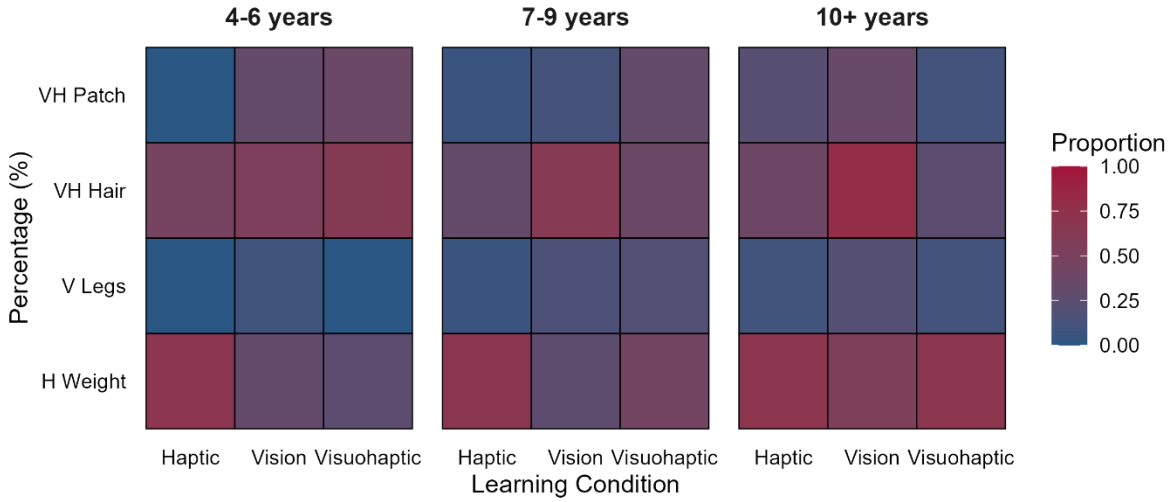


Supplemental 10: Qualitative analysis

In a qualitative assessment of category representations all participants were asked to name the features that differentiated the 'Neems' from the 'Dorps' (or vice versa, counterbalanced across participants) after the visuohaptic categorisation task. We coded their answers into four different options, each based on the feature modality as follows: Visual only feature correct (leg colour) or incorrect (body colour etc.); haptic-only feature correct (weight) or incorrect (compressibility); visuohaptic features correct (patches or hair type) or incorrect (body or leg shape), yielding a total of 8 options. We also added a further option in which the child demonstrated that they were guessing. The total number of explicitly reported diagnostic features, across the age groups and learning modalities are represented in Figure S3. We found that the feature most frequently reported by the children who learned the objects through touch was the haptic-only feature, followed by the visuohaptic feature. The children who learned the objects in the visual-only condition, mainly reported a visuohaptic feature, followed by the haptic and visual features. The children who learned the objects visuohaptically reported fewer features overall, although the haptic feature was more frequently reported, followed by a visuohaptic feature. These qualitative responses were aligned with our accuracy analysis of the cross-category objects where the haptic cue weight followed by the visuohaptic cue hair affected participant categorisation responses compared to the patch and leg colour cues.

Figure S8

Cues Identified as being diagnostic of category membership across age and learning modality conditions



Note. Proportion of Participants who reported using each diagnostic feature is expressed per age group per learning condition from Blue (0) to Red (1).

Supplemental 11: Stimulus Properties

Table S6

Generalisation Stimuli: feature characteristics of objects with a cross-category features

		Diagnostic Features			
Stimulus	Category	Leg Colour	Body Colour	Patches	Button
1	Neem	orange	red	mesh	large
2	Neem	purple	light blue	hessian	large
3	Neem	dark blue	red	ribbon	large
4	Neem	purple	red	hessian	small
5	Neem	yellow	orange	mesh	larger
6	Neem	dark blue	dark blue	mesh	larger
7	Neem	dark blue	orange	felt	larger
8	Neem	purple	orange	hessian	smaller
9	Dorp	dark blue	dark blue	ribbon	smaller
10	Dorp	yellow	red	felt	smaller
11	Dorp	yellow	dark blue	mesh	smaller
12	Dorp	orange	dark blue	ribbon	large
13	Dorp	purple	light blue	felt	small
14	Dorp	orange	orange	ribbon	small
15	Dorp	yellow	light blue	hessian	small
16	Dorp	orange	light blue	felt	larger

Note. Cells with grey shading are the cross-category features.

Supplemental 12: Stimulus Creation

Stimuli were initially photographed using a Nikon camera at a distance of 30cm. These objects were rotated 25° to the right in order to ensure all object features were in frame. See A for photographing set up:

Figure S9

Stimulus Photography set up

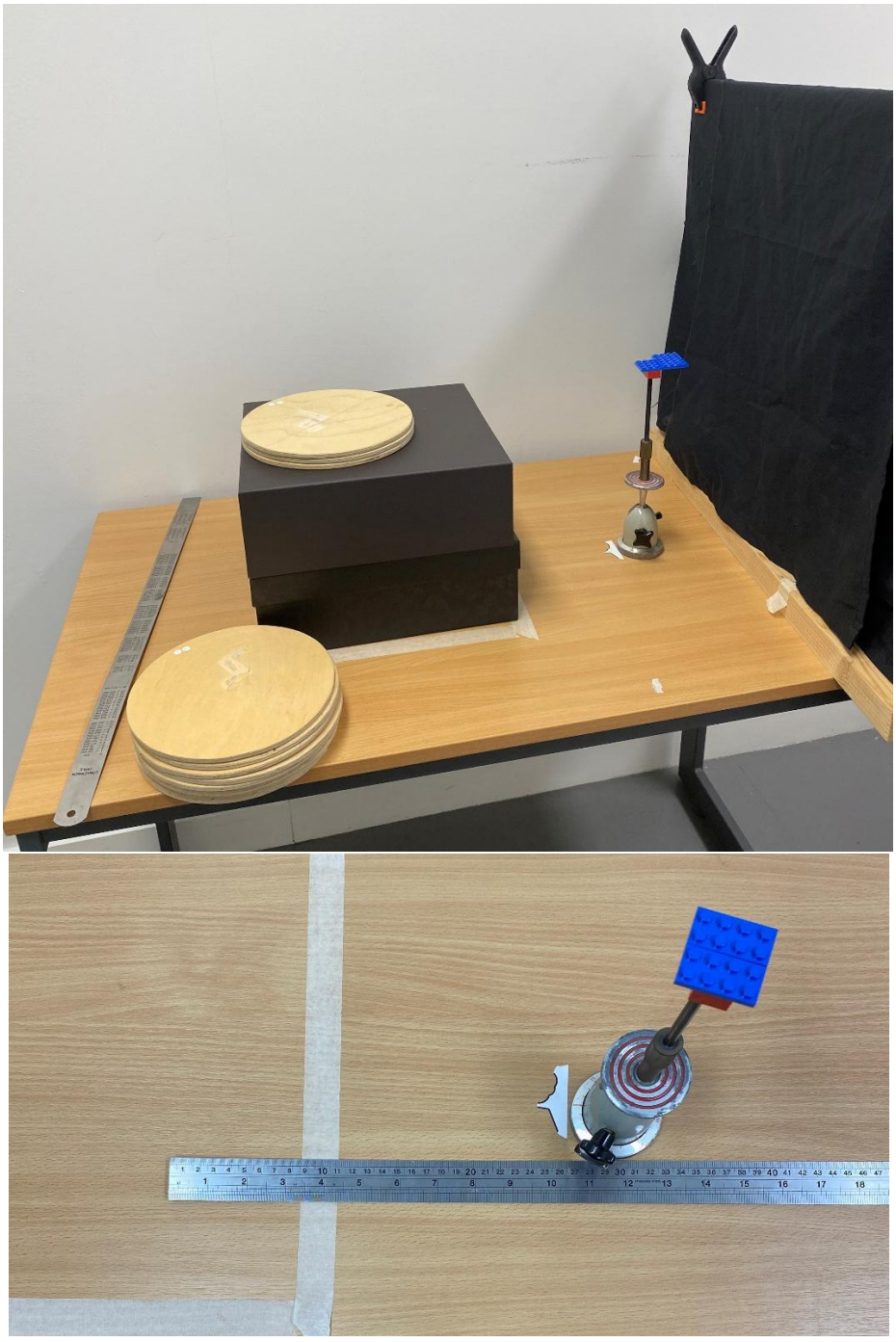
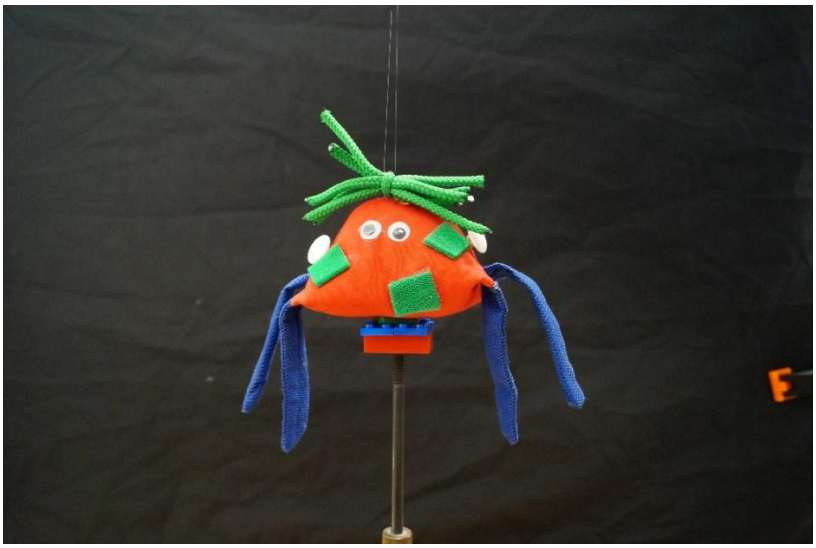


Figure S10

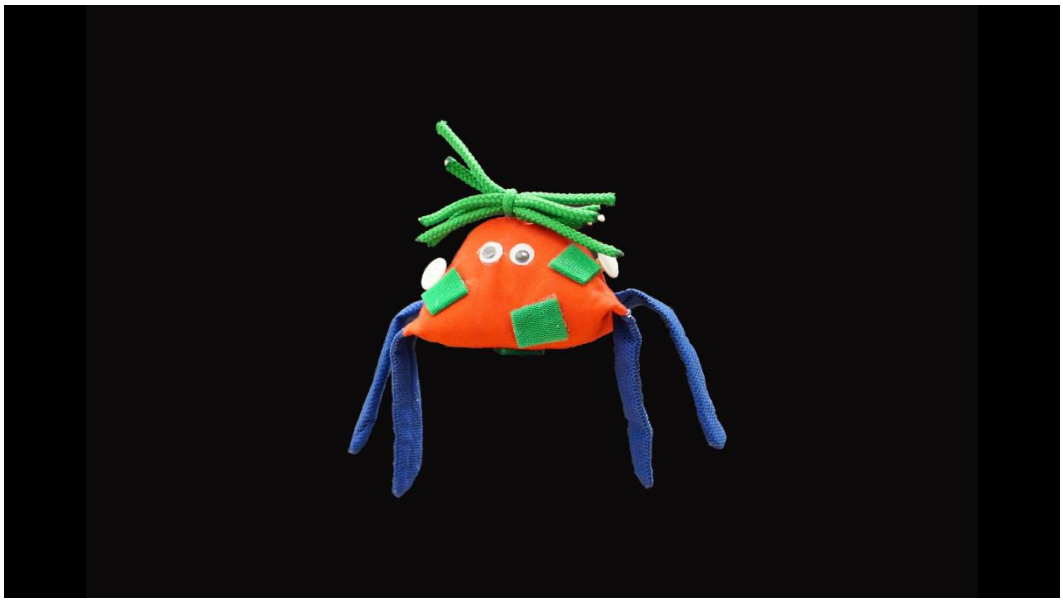
Initial Stimulus photograph



All background information was subsequently removed, and the dimensions of the image were fixed at 1980 by 1080 pixels to ensure all stimuli matched our display screen dimensions.

Figure S11

Stage 2 image processing output



We subsequently completed some post processing to increase the saliency of the diagnostic feature of patch:

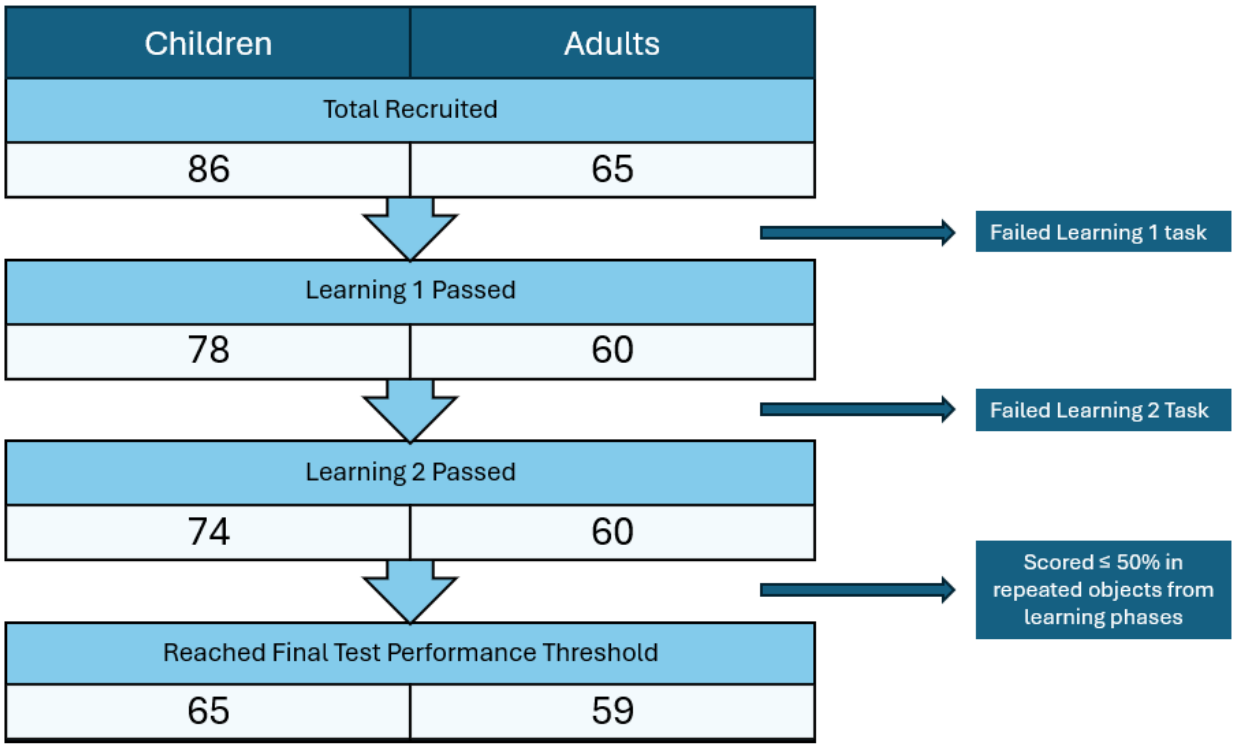
Figure S12

Final Stimulus utilised in visual learning and test phases

Supplemental 13: Participant Exclusions

Figure S13

Participant count at each experimental phase



Supplemental 14: Proportion of Fixations to AOIs -size of AOI as a covariate

Areas of interest (AOIs) varied in physical size across stimuli, raising the possibility that larger AOIs might receive a greater proportion of fixations simply by chance. To adjust for these differences, we included the log-transformed and mean-centered AOI area (in cm^2) as a covariate in our statistical model in order to assess if this may be the cause of the observed effects of AOI type (Visual; Visuohaptic; Other Body Features). This approach statistically controls for the expected influence of AOI size on fixation proportions, rather than imposing a fixed reweighting. This method is consistent with recommended practices in eye-tracking research when AOIs differ in extent (e.g., Holmqvist et al., 2011; Orquin & Holmqvist, 2018), and avoids the strong assumptions implied by simply dividing fixation proportions by AOI area.

We conducted a linear mixed-effects model to examine the proportion of fixations directed to each AOI type as a function of age group (child vs. adult) and AOI type (visual, visuohaptic, other body features), while controlling for differences in AOI size. To account for variation in AOI size across stimuli, we included log-transformed and AOI-type centered AOI area as a covariate. AOI type was treated as a within-subjects factor, and age group as a between-subjects factor. The model included random intercepts for participant and for AOI type nested within participant, to appropriately model the repeated-measures structure (i.e., multiple AOI types per participant). In total, the model included 1,402 observations from 138 participants. The model was fitted using REML with the Nelder–Mead optimizer.

The model explained a moderate amount of variance, with a conditional $R^2 = 0.34$ and marginal $R^2 = 0.15$, indicating that fixed effects explained 15% of the variance in fixation proportions, and including random effects explained 34% in total. The model

residuals passed visual inspection, and multicollinearity was low after centring AOI area (VIFs < 4.6).

There was a significant main effect of age group, $F(1, 96.14) = 5.94$, $p = .017$, partial $\eta^2 = .06$, indicating overall differences in fixation proportions between adults and children. Overall, children allocated a greater proportion of fixations ($M = 0.64$, $SD = 0.48$, 95% CI [0.61, 0.68]) than adults ($M = 0.58$, $SD = 0.51$, 95% CI [0.55, 0.62]), averaged across AOI types and controlling for AOI size.

There was also a main effect of AOI type [$F(2, 187.95) = 95.79$, $p < .001$, partial $\eta^2 = .50$], after statistically adjusting for AOI size. Post-hoc tests were conducted using estimated marginal means (emmeans), Bonferroni-adjusted for multiple comparisons. Collapsing across age groups, visual AOIs received a significantly higher proportion of fixations ($M = 0.71$, $SD = 0.40$, 95% CI [0.68, 0.71]) than either visuohaptic AOIs ($M = 0.44$, $SD = 0.48$, 95% CI [0.41, 0.48];[$t(187) = 13.87$, $p < .0001$]) or other body features ($M = 0.60$, $SD = 0.36$, 95% CI [0.56, 0.63]; [$t(175) = 6.17$, $p < .0001$]). Participants also fixated less on visuohaptic AOIs than on other body features ($t(229) = -7.37$, $p < .0001$).

There was also a significant main effect of AOI size [$F(1, 1239.01) = 14.57$, $p < .001$, partial $\eta^2 = .01$], showing that larger AOIs attracted a greater proportion of fixations overall. The interaction between age group and AOI type was not significant [$F(2, 188.02) = 2.19$, $p = .114$].

Supplemental 15: Duration of Fixations to AOIs- size of AOI as a covariate

A linear mixed effects model was conducted to assess the potential differences in mean fixation duration within each of the AOI regions compared across age groups (child versus adult), while controlling for AOI size. To account for repeated measures, random intercepts were included for participants, with AOI type nested within participants. The model was fitted using maximum likelihood estimation on log-transformed fixation durations to correct for positive skew. AOI size (log-transformed and mean-centered within AOI type) was included as a covariate to control for systematic size differences between AOIs.

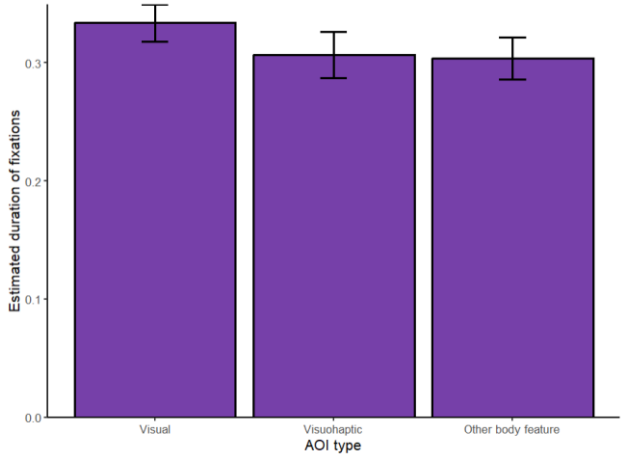
The model explained 18.9% of the total variance (conditional $R^2 = 0.189$) and 1.5% of the variance in fixed effects alone (marginal $R^2 = 0.015$). Model diagnostics showed no violations of assumptions, and checks for multicollinearity indicated acceptable VIFs for all predictors (all adj. VIFs ≤ 2.09). The Type III ANOVA revealed a main effect of AOI type [$F(2, 155.70) = 8.57$, $p < .001$, $\eta_p^2 = .10$]. Bonferroni-adjusted pairwise comparisons between AOI types showed that fixations to visual AOIs ($M = 0.33$, $SD = 0.16$, 95% CI [0.32, 0.35]) were significantly longer than fixations to visuohaptic AOIs ($M = 0.31$, $SD = 0.15$, 95% CI [0.29, 0.33]; $t(190) = 2.70$, $p = .023$) and other body features ($M = 0.30$, $SD = 0.14$, 95% CI [0.29, 0.32]; $t(178) = 3.19$, $p = .005$). There was no significant difference between visuohaptic AOIs and other body features ($p = 1.000$). These effects are visualized in Figure SX, which shows estimated marginal means of fixation duration by AOI type, adjusted for AOI size. Error bars represent 95% confidence intervals.

There was also a significant effect of AOI size [$F(1, 1246.47) = 5.71$, $p = .017$, $\eta_p^2 = .005$]. The main effect of age group was not significant [$F(1, 104.58) = 0.04$, $p = .84$, $\eta_p^2 < .001$], nor was the interaction between age group and AOI type [$F(2, 155.78) = 0.16$, $p = .85$, $\eta_p^2 = .002$]. No significant interaction effects were observed. This indicates that age

group did not modulate fixation durations across AOI types, and the observed differences in fixation duration were driven primarily by AOI type, independent of AOI size and age.

Figure S14

Mean fixation duration within each AOI type



Note. Error bars represent 95% CIs