Growing Minds, Integrating Senses: Neural and Computational Insights into Age-related Changes in Audio-Visual and Tactile-Visual Learning in Children

Nina Raduner[a,b,c,d,f], Carmen Providoli[a,c,d,e], Sarah Di Pietro[a,e], Maya Schneebeli[a,e], Iliana I. Karipidis[a,d], Ella Casimiro[e,f], Saurabh Bedi[e,f], Michael Von Rhein[c,e], Nora Maria Raschle[b,d,e], Christian C. Ruff[d,e,f], Silvia Brem[a,d,e]

Affiliations:

[a] Department of Child and Adolescent Psychiatry and Psychotherapy, University Hospital of Psychiatry, University of Zurich, Zurich, Switzerland

[b] Jacobs Center for Productive Youth Development, University of Zurich, Zurich, Switzerland

[c] Child Development Center, University Children's Hospital Zurich, Zurich, Switzerland

[d] Neuroscience Center Zurich, University and ETH Zurich, Zurich, Switzerland

[e] URPP Adaptive Brain Circuits in Development and Learning (AdaBD), University of Zurich, Zurich, Switzerland

[f] Department of Economics, Zurich Center for Neuroeconomics, University of Zurich, Zurich, Switzerland

Corresponding author

Prof. Dr. Silvia Brem
Neumünsterallee 9
8032 Zürich
Switzerland
silvia.brem@uzh.ch

Abstract

Multisensory processing and learning shape cognitive and language development, influencing how we perceive and interact with the world from an early age. While multisensory processes mature into adolescence, it remains poorly understood how age influences multisensory associative learning. This study investigated age-related effects on multisensory processing and learning during audio-visual and tactile-visual learning in 67 children (5.7–13 years) by integrating behavioural and neuroimaging data with computational methods. A reward-learning drift diffusion model revealed that older children processed information faster and made more efficient decisions on multisensory associations. These age-related increases coincided with higher activity in brain regions associated with cognitive control, multisensory integration, and memory retrieval, specifically during audio-visual learning. Notably, the anterior insula exhibited heightened activation in response to lower reward prediction errors, indicative of increased sensitivity to negative feedback with development. Finally, reward prediction errors and values modulated activation in reward processing and cognitive control regions, with this modulation remaining modality-independent and largely stable across age. In conclusion, while children employ similar learning strategies, older children make decisions more efficiently and engage neural resources more strongly. Our findings reflect ongoing maturation of neural networks supporting multisensory learning in middle childhood, enabling more adaptive learning in later childhood.

Keywords

multisensory learning
audio-visual
tactile-visual
reward prediction error
computational modelling
fMRI
children

Highlights

- faster information processing in older children during a multisensory learning task
- increasing brain activation with age in visual, parietal, and frontal regions
- learning rate related to middle frontal and anterior cingulate cortex activity
- value and reward prediction error processing are independent of age and modality
- heightened response to negative reward prediction errors with increasing age

1 Introduction

Efficient processing and integration of sensory information from multiple modalities including vision, hearing, and touch, are essential for interacting with our surroundings and adapting to the complexities of our environment. On a daily basis, humans encounter numerous stimuli by various senses, requiring appropriate processing and integration to be accurately understood (Van Atteveldt et al., 2014). Multisensory processing describes the perception of stimuli in multisensory environments (Murray et al., 2016). Multisensory learning (MSL) is the process of acquiring associations between stimuli from different senses (Lauzon et al., 2022), while multisensory integration (MSI) occurs when the brain combines information from multiple senses to form a unified percept (Stein et al., 2020; Wallace et al., 2020). When stimuli from different modalities provide congruent information, they are integrated and perceived as originating from the same source. Research using electrophysiology, fMRI, and behavioural studies has shown that successful associative MSL leads to MSI (Butler et al., 2011; Lauzon et al., 2022). Since these processes are sequential (multisensory processing is needed for MSL which is also the basis for MSI), they are often studied in one framework. MSI is particularly significant in naturalistic learning environments and classrooms, where multisensory exposure is common (Barutchu et al., 2019). Moreover, MSI skills are associated with enhanced cognitive processes including memory performance and reading abilities (Barutchu et al., 2011, 2019; Birch & Belmont, 1965; Denervaud et al., 2020; Dionne-Dostie et al., 2015). MSI likely promotes learning by enabling the effective combination of sensory inputs for better encoding, storage, and retrieval of information. For example, MSI underpins language learning, production, and development by integrating auditory and visual information like sounds and lip movements (Wallace et al., 2020).

Given the importance of MSL and MSI in cognitive development, it is crucial to investigate MSL in childhood and how it is similar or different across different sensory modalities. However, most studies on MSL have focused on adults and the studies investigating MSL in children mainly concentrated on audio-visual speech sound mapping (e.g., Frei et al., 2025; Karipidis et al., 2021; Wang et al., 2020). Only a few investigations targeted the learning and processing of non-linguistic audio-visual integration (e.g., Altarelli et al., 2020; Batson et al., 2011; Pasqualotto et al., 2024; Seitz et al., 2007). Moreover, while MSL can encompass signals from all senses, minimal attention has been directed towards the interplay between touch and vision or any other combination of stimulus modalities (see meta-analysis from Gao et al., 2023).

Despite the growing interest in and the importance of MSL, relatively few studies have examined its neural underpinnings (e.g., Butler et al., 2011; Lauzon et al., 2022), with most research focusing on MSI or failing to distinguish between the two processes. The meta-analysis by Gao et al. (2023) further highlights that audio-visual integration has been widely studied, whereas only a few studies examined tactile-visual integration (such as Gentile et al., (2013, 2015); Limanowski & Blankenburg, (2016, 2017)). In the mentioned meta-analysis, audio-visual integration was associated with changes in neural activation of early sensory regions, subcortical areas, and higher association areas supporting a flexible neural pathway model for audio-visual integration. Such a model is supported by research linking MSI and learning to an extended cortico-striato-thalamic network (Lakatos et al., 2007; Murray et al., 2016; Tyll et al., 2011), including early sensory regions (Ghazanfar & Schroeder, 2006; Lakatos et al., 2007; Noesselt et al., 2007; Stein & Stanford, 2008). Moreover, reviews summarising brain regions involved in MSI identified similar regions to be relevant for these processes: the auditory cortex, visual cortex, somatosensory cortex, superior temporal cortex, and frontal

regions (Murray et al., 2016; Richlan, 2019). Exploratory meta-analyses on the few existing studies on tactile-visual integration reveal that inferior temporal areas, the fusiform gyrus, and inferior parietal areas are associated with tactile-visual integration (Gao et al., 2023). This highlights the need for further research on the brain areas underlying MSL, particularly during childhood, when these abilities are still developing.

Although infants already demonstrate the ability to process and integrate multisensory information (Patterson & Werker, 1999, 2003), a deeper understanding of how these abilities develop over childhood is crucial for advancing learning research. MSI has not yet reached full maturity in children younger than eight years old and maturation of MSI skill continues into adolescence (Desjardins et al., 1997; Dionne-Dostie et al., 2015; Gori et al., 2008; Nardini et al., 2008). Research suggests the temporal window for audio-visual integration is still immature in 11-year-old children (Dionne-Dostie et al., 2015). According to Murray et al. (2016), children's MSI abilities are initially more reliant on low-level stimulus feature processing but gradually shift toward higher stimulus feature processing and therefore cortical involvement. This process reflects the brain's developmental trajectory where subcortical regions mature before cortical regions. This implies that younger children, who depend more on the processing of low-level features through subcortical and primary sensory areas, might have limited MSI capabilities for complex learning tasks requiring greater levels of cognitive control. With children's increasing age, however, the integration of sensory information progressively involves higher-order brain regions, such as the frontal cortex. This in turn allows for more flexible and sophisticated MSI that is crucial for advanced learning, problem-solving, and adaptation in dynamic environments. As a result, these maturation processes allow for more complex multisensory learning and integration as children grow (Murray et al., 2016).

To better understand the developmental processes, computational modelling provides a powerful tool for examining the cognitive mechanisms underlying MSL, deepening the neural and developmental insights. One approach in computational modelling integrates a learning process, reflecting trial-wise belief updating, and a choice process, translating computational quantities (e.g., values) into behavioural or neural responses (Mathys et al., 2011). This study employs a combined reinforcement learning model as the learning process and drift diffusion model as the choice process (RLDDM; Pedersen et al., 2017) to investigate multisensory processing and learning in children (see also Fraga-González et al., 2025; Frei et al., 2025). Reinforcement learning, one of the core mechanisms for feedback learning, compares observed outcomes (rewards) with expected outcomes (values) and adjusts expectations based on the discrepancy – known as the reward prediction error (RPE; Niv & Schoenbaum, 2008; Subramanian et al., 2022). The Rescorla-Wagner (RW) model describes this process by weighting the RPE with the learning rate to update values (Rescorla & Wagner, 1972). On the other hand, the drift diffusion model quantifies the decision-making process through the drift rate, decision boundary, and non-decision time (Pedersen et al., 2017; Ratcliff, 1978; Ratcliff & Rouder, 1998). By combining these models, we can identify latent cognitive processes like value updating, learning rates, and RPEs, and relate them to the brain activation during learning.

Building on these computational frameworks, previous research has identified specific neural correlates of the cognitive parameters involved in learning. RPEs are consistently linked to stronger brain responses in the striatum, the medial frontal cortex, and hippocampus reflecting their relation to learning and memory (Cohen & Ranganath, 2005; Fraga-González et al., 2025; Garrison et al., 2013; McClure et al., 2003; O'Doherty et al., 2003; Pagnoni et al., 2002). Similarly, value processing involves regions such as the ventral striatum, anterior cingulate cortex, anterior insula, medial orbitofrontal cortex, supplementary motor area, and

inferior parietal lobe (Liu et al., 2011; R. P. Wilson et al., 2018). These regions are pivotal in integrating motivational and emotional aspects of decision-making (Liu et al., 2011). The learning rate has been linked to activation in the inferior frontal gyrus, anterior insula, dorsal anterior cingulate cortex, and ventral caudate (Chien et al., 2016; O'Reilly, 2013; Wu et al., 2017). By mapping neural correlates of computational parameters such as values, reward prediction errors, and learning rates, our study aims to computationally characterize the perceptual and learning processes underlying MSL in children and study their change during development.

To further explore these neural and cognitive processes, the current study investigated MSL in middle childhood, focusing on learning sensory associations of different modalities (audio-visual vs. tactile-visual) and on age-related changes therein. The primary aims were: 1) to **investigate modality-specific differences in children's multisensory processing and learning** by comparing audio-visual (AV) and tactile-visual (TV) associations, using computational modelling of behavioural performance and neural activity data, 2) to characterize the **age-related changes in MSL** by analysing age-related differences in processing speed, decision-making and associated neural activity, and 3) **to examine how children process feedback during MSL**, specifically looking at age-effects on RPE and value-based learning for audio-visual and tactile-visual learning. We hypothesised that younger children would rely more on subcortical and primary sensory regions, while older children would engage higher-order cortical areas during MSL, indicating the increased employment of neural resources (Murray et al., 2016). Accordingly, we anticipated neural activation in brain regions previously associated with value processing (Liu et al., 2011; R. P. Wilson et al., 2018), reward prediction error processing (Cohen & Ranganath, 2007; Fraga-González et al., 2025; Garrison et al., 2013; McClure et al., 2003; Pagnoni et al., 2002), and learning rate processing (Chien et al., 2016; O'Reilly, 2013; Wu et al., 2017). We expected potential age-dependent variations likely to be pronounced in higher-cortical areas such as prefrontal and cingulate regions (Yaple et al., 2020). With this study, we bridge the gap in the existing literature by investigating non-linguistic audio-visual and tactile-visual associative learning across middle childhood.

We employed a cross-sectional design with 67 children (aged 5.7–13.0 years) completing a multisensory learning task involving AV and TV stimuli. Computational modelling, incorporating reinforcement learning and drift diffusion frameworks, was applied to characterise the cognitive processes underpinning learning. Functional MRI was employed to pinpoint brain regions engaged in modality-general and modality-specific learning processes and to characterize age-related changes in the activation of involved brain networks. Parametric modulation of the BOLD signal with model parameters finally allowed us to examine the neural correlates underlying specific learning and decision processes, as well as potential age-related changes in the neural processing.

2 Material and methods

2.1 Participants

This study included 80 healthy, native (Swiss-)German-speaking children. Children with either neurological or psychiatric disorders or intellectual disabilities (except for specific learning disorders or ADHD) were not included to participate in this study. Following the exclusion of one subject due to technical problems during scanning, another two subjects had to be excluded due to excessive movement in all runs (see chapter 2.3) and a high number of omissions (see Supplementary Material A.6). Additionally, 10 children were excluded from the final analysis as they had only one AV or TV run with sufficient MR data quality (no excessive movement). Consequently, the final sample comprised 67 children (*M* = 8.35 ± 1.82 years, 36 girls) with at least one high-quality run in both the AV and TV modalities. There was no significant age difference between girls and boys (see Supplementary Material A.1). All children had estimated IQ scores above 80. All participants had normal or corrected to normal vision and normal hearing. Mean IQ, processing speed, and attention scores were in the normal range (see **Table 1**), indicating age-appropriate cognitive abilities for all subjects (see Supplementary Material A.1 for more information). A total of ten children had either a clinical diagnosis (attention deficit hyperactivity disorder, developmental dyslexia, or developmental language delay) and/or clinically significant values in the CBCL questionnaire (Döpfner et al., 2014; see Supplementary Material A.1 for more information). The following analyses were conducted with and without these ten participants (see Supplementary Material B.1). This approach ensures that the results are not biased by potential neurodevelopmental differences while also allowing for a more comprehensive analysis that includes all participants.

Oral consent was obtained from all children, while written consent was provided by a parent or guardian for all children. The study was approved by the local ethics committee of the Canton of Zurich and neighbouring cantons in Switzerland (BASEC-Nr: 2022-01368), and the participants received vouchers and gifts in appreciation of their participation.

Table 1

Demographic characteristics of study participants

	M (*SD*) [range]
Age [y]	8.35 (1.82) [5.7-13]
Sex [f:m]	36:31
Handedness[a] [l:a:r]	6:3:57
IQ (verbal and non-verbal)	106.5 (10.6)
Rapid automatized naming [T-value]	50.1 (7.45)
Processing speed [T-value]	50.12 (7.97)
Sustained attention median RT[b] [PR]	51.55 (29.43)

Note. The handedness of the subjects was determined using the Edinburgh Handedness Scale (Oldfield, 1971). IQ values were estimated using the Peabody Picture Vocabulary Test – fourth edition (PPVT-IV; Lenhard et al., 2015) and subtests from the Wechsler Intelligence Scale for Children – fifth edition (WISC-V; Wechsler, 2017) or the Wechsler Preschool and Primary Scale of Intelligence – fourth edition (WPPSI-IV; Wechsler, 2012). Rapid automatized naming and processing speed measured with IDS-2 (Grob & Hagmann-von Arx, 2018). Sustained attention measured with KiTAP (Zimmermann et al., 2002). [a] One child's data are missing. [b] Percentiles are missing for two children due to the absence of normative values for their age. f = female, m = male, l = left, a = ambidextrous, r = right, RT = reaction time, PR = percentiles.

2.2 Experimental design

During the neuroimaging session, participants completed an audio-visual and tactile-visual multisensory learning task (adapted from (Frei et al., 2025)) embedded in a child-friendly story in the MRI scanner. The multisensory learning task is a two-alternative forced choice task where the participants had to learn multisensory associations between visual and auditory or visual and tactile stimuli based on feedback. Participants were presented with two symbols alongside either a sound or vibration pattern, representing a rocket's sound or engine trembling, respectively. Participants guessed which rocket the sound or trembling belonged to by pressing a corresponding button. They were told to learn associations over time through feedback, which was implicitly suggested to be probabilistic. The task was presented using goggles (CinemaVision 20/20, Resonance Technology, Northridge, CA), MR-compatible over-ear headphones (MR confon GmbH, Magdeburg, Germany, http://www.mr-confon.de/de/), a tactile stimulation device (mini-PTS piezoelectric stimulator device from dancer design, http://www.dancerdesign.co.uk/) stimulating the right pinkie, and the PsychoPy® software (Version 2022.1.1, Peirce et al. (2019); https://www.psychopy.org/). To ensure that each child could perceive the stimuli clearly during the MR session, the sound volume was adjusted, and vision was corrected individually as needed.

Not all children were able to complete every task according to the experimental protocol and therefore runs had to be repeated or excluded due to excessive movement, lack of motivation (too many omissions, see Supplementary Material A.6), or fatigue. For the purposes of this analysis, only data from children who completed at least one audio-visual and one tactile-visual run with adequate data quality were included (18 children completed two runs of each modality, 39 completed one run of each, 5 completed two AV and one TV run, and 5 completed one AV and two TV runs).

As depicted in **Figure 1** during the inter-trial interval (ITI), the task background consisted of a space image (dark background with stars), two blank rockets on either side of a bigger star in the middle acting as the fixation cross. These elements were constantly displayed during the whole task. During stimulus presentation, one auditory stimulus was presented simultaneously with two visual stimuli appearing on either rocket for a duration of 2 seconds (**Figure 1**). In TV runs, a tactile stimulation was presented instead of an auditory sound (see Supplementary Material A.2). The participants were then instructed to choose the matching symbol by pressing the left or right button on an MR-compatible two-button response box for the left or right choice, respectively, and their choice was indicated by a blue frame during the inter-stimulus interval (ISI). The participant's choice was recorded during stimulus presentation and the ISI. After the response, they received feedback on their choice for 1.8s, which was followed by an ITI before the next trial began. In most trials, feedback was positive after a correct choice, negative after an incorrect choice, or neutral after no choice was observed (see Supplementary Figure A.2B). In probabilistic trials, participants received incorrect feedback (e.g., positive feedback after a wrong choice or vice-versa; see Supplementary Material A.3 for more information). The ITI and ISI were gamma distributed and had a mean duration of 1.87s [0.76 – 3.45s] and 3.45s [2.28 – 4.83s], respectively. After the first six participants, we slightly changed the timing of the ISI and ITI to provide more decision time. For these participants, the mean durations were 2.06s [0.79 – 3.39s] and 2.67s [1.11 – 4.34s], respectively for the ITI and ISI. Each run consisted of 44 trials and total task duration was 401 seconds (384 seconds for the first 6 subjects).

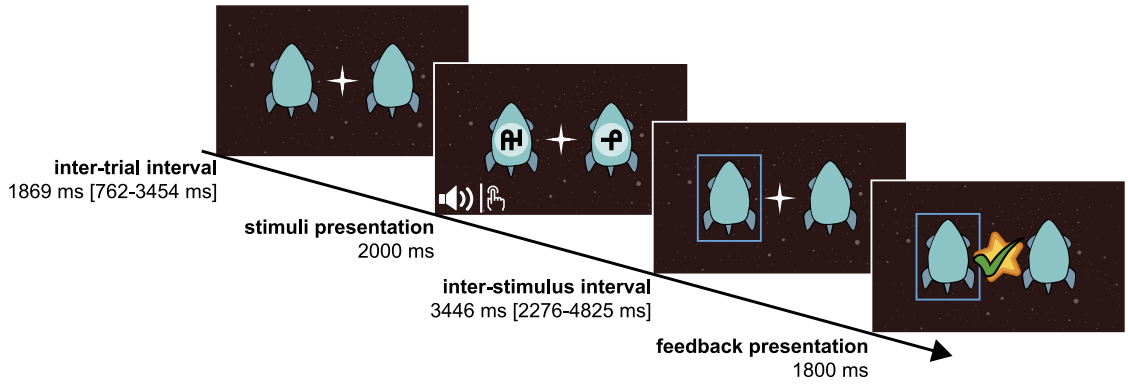
In total, there were four visual stimuli and four auditory or tactile stimuli per run, resulting in 16 possible combinations of which four were matching combinations. For the MR runs, we prepared six sets of visual, three sets of auditory stimuli, and three sets of tactile stimuli (see Supplementary Material A.2). As visual stimuli, artificial symbols from the PseudoSloan font

were used (Vildavski et al., 2021). The pseudo letters were chosen by the complexity scores provided by Vildavski et al. (2021). Each set was matched such that it included one stimulus with low (M = 3.55, SD = 0.02), two stimuli with medium (M_1 = 4.27, SD_1 = 0.09 and M_2 = 4.49, SD_2 = 0.05) and one stimulus with high complexity (M = 5.16, SD = 0.05). Additionally, we also balanced the selection of pseudo letters between sets. For the audio-visual run, environmental sounds were used as auditory stimuli, which were selected from the mixkit sound effects database (https://mixkit.co/free-sound-effects/). Here, we chose sounds that could not be easily identified or named. After selection, the sounds were cut to a 2 second duration and normalized using Audacity® recording and editing software (version 3.5.1; Team Audacity, 2024). The mean Shannon spectral entropy (Cover & Thomas, 2005; Misra et al., 2004; Obin & Liuni, 2012), a measure of sound complexity, was calculated for each stimulus using the spectralEntropy function (Signal Processing Toolbox, version 9.1) in MATLAB (version R2022b; MathWorks Inc, 2022). Each set of environmental sounds consisted of one sound with a low (M = .327, SD = 0.03), two with a medium (M_1 = .476, SD_1 = 0.05 and M_2 = .564, SD_2 = 0.04), and one with a high entropy value (M = .705, SD = 0.04). Lastly, we created three different sets for the tactile stimulation. Each set of tactile stimuli consisted of four different vibration patterns lasting for 2 seconds. Each set contained two stimulations with distinct stable frequencies (low complexity), one was increasing or decreasing (medium complexity), and one had an (inverted) u-shape (high complexity; see Supplementary Figure A.2C). Ensuring that the stimuli varied based on complexity within a set was done to increase the discriminability between stimuli. The following variables were pseudo-randomised for all participants: the order of tasks (AV and TV) in the MR scanner, the trial structures (the sequence of stimulus presentation), the visual, auditory, and tactile sets used, and the assignment of the stimuli to the four correct combinations for each run.

Each participant completed behavioural runs (one for each modality) in a behavioural assessment session 1 – 42 days (M = 10.6 days, SD = 8.1 days) before the MR session (pre runs). The performance from the pre runs was then used to adapt the task difficulty in the MR scanner to ensure that the task was neither too easy nor too difficult. Task difficulty was either adapted by the trial structure or by probabilistic feedback trials. Additionally, some of the stimuli triplet combinations were presented more often than others (see Supplementary Material A.3 for more information about trial structure, probabilistic feedback, and stimulation frequency).

Figure 1

Multisensory learning task



Note. Illustration of the multisensory learning task showing one trial for the audio-visual (AV) and tactile-visual (TV) task with presentation durations. Two rockets, each displaying a symbol, were presented, while either a sound (AV) or a vibration pattern (TV) was played. Children then determined which rocket the sound or vibration belonged to by button press. Their chosen rocket

was highlighted with a blue frame. Feedback was afterwards provided via a yellow star in the center of the screen, accompanied by a checkmark for correct choices. Incorrect selections were indicated by a gray star and a red cross. After omissions, a double sided, yellow arrow was displayed, indicating no choice, and reminding the participants to response faster in future trials. See supplementary material A.2 for a depiction of all feedback stimuli.

2.3 MRI data acquisition and preprocessing

MRI data was acquired on a Philips Achieva 3 Tesla scanner (Best, The Netherlands) using a 32-channel coil. T2* weighted functional images were acquired with a whole-brain gradient echoplanar image (EPI) sequence (five dummy scans followed by 291 dynamic scans, repetition time TR = 1.395 s, echo time TE = 35 ms, 44 slices, voxel size = 3.0 x 3.0 x 3.0 mm3, slice gap = 0.3 mm, matrix size 64 x 62 px, flip angle = 80°, multiband factor = 2, sofTone factor = 2, sensitivity-encoding (SENSE) reduction factor = 2). Due to changes in the timing of the task (see chapter 2.2), the EPI sequence of 6 participants only included 278 dynamic scans while all the other parameters were the same. T1-weighted anatomical images were obtained for co-registration using a magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence (TR = 6.96 ms, TE = 3.2 ms, aligned at AC-PC, flip angle = 9°, voxel size = 1.0 x 1.0 x 1.0 mm^3, field of view = 270x255 mm^2, number of slices = 176).

MATLAB (version R2020b, MathWorks Inc, 2020) toolbox SPM12 (7219, Penny et al., 2007) was used for preprocessing and whole-brain analysis, which included slice-time correction, realignment, co-registration, and segmentation. Using the Template-O-Matic toolbox (Wilke et al., 2008), a paediatric template was created for normalization (age matched structural data). Voxels were resampled to 3 x 3 x 3 mm^3 isotropic voxels. An 8-mm full width at half maximum (FWHM) Gaussian kernel was applied for smoothing. After preprocessing, we repaired volumes with a composite scan-to-scan motion above 1.5mm (Euclidean distance with a radius of 65 mm) with the ArtRepair toolbox (Mazaika et al., 2007), using linear interpolation between the nearest unrepaired scans. After this, we flagged volumes with and without movement surrounded by scans with motion above 1.5mm. Runs with more than 20% of flagged volumes were excluded from the final analyses, resulting in a total of 18 runs being omitted (movement was not correlated with age; see Supplementary Material B.9). Whenever possible, we attempted to repeat runs with excessive movement during the MR session. Flagged scans were accounted for by including one binary regressor of no interest for each flagged volume (scrubbing).

2.4 Computational modelling of values, RPEs, and learning rate

For the learning process, we implemented the trial-wise updating mechanism using a Rescorla-Wagner (RW) model which is an established model for reinforcement learning (Rescorla & Wagner, 1972). To model basic multisensory processing involved in constructing the computational quantities of the mental model and the choice processes that translate them into responses, a drift diffusion model (DDM; Ratcliff, 1978) was implemented based on the approach outlined by Pedersen et al. (2017). This generative model allowed us to estimate parameters of the trial-wise learning process, inferred from the accuracy and reaction times of the responses from each trial (Pedersen et al., 2017). This approach permitted the calculation of trial-wise values for each multisensory association, drift rates, and RPEs.

The model used in this study is illustrated in **Figure 2**. The RW model is used to calculate the RPEs and update the values V associated with the responses to the current multisensory association presented in each trial. The RPE for a trial t is defined as the difference between the actual reward R and the expected reward (= value V):

$$RPE_t = (R_t - V_t) \tag{1}$$

Then, the value V for each trial t was updated by the RPE weighted by the learning rate η:

$$V_t = V_{t-1} + \eta * (R_{t-1} - V_{t-1}) \tag{2}$$

Therefore, the learning rate is a weighting factor that indicates the extent to which values are updated after each trial (for a more detailed interpretation, see Zhang et al. (2020)). In this study, only the value for the chosen response to the specific multisensory stimulus combination was updated.

Then for each trial t, the values for the left $\bar{V}_{left,t}$ and right $\bar{V}_{right,t}$ stimulus combination were normalised to ensure that the sum of both values was 1:

$$\bar{V}_{left,t} = \frac{V_{left,t}}{(V_{left,t} + V_{right,t})}, \bar{V}_{right,t} = 1 - \bar{V}_{left,t} \tag{3}$$

Theses normalised values from the RW were then entered in the DDM to calculate the drift rate v by multiplying the difference between the left and right value with the drift weight v_{mod}:

$$v_t = \left(\bar{V}_{left,t} - \bar{V}_{right,t}\right) * v_{mod} \tag{4}$$

Therefore, the drift weight is a scaling factor and corresponds to the sensitivity to value differences (e.g., with a high drift weight even a small difference between the values leads to high drift rate; Pedersen et al. (2017)). We used the wiener first passage time (WFPT) algorithm (Navarro & Fuss, 2009) to fit the choice behaviour and reaction time data:

$$RT(x) \sim WFPT[a, \tau, z, v(t)] \tag{5}$$

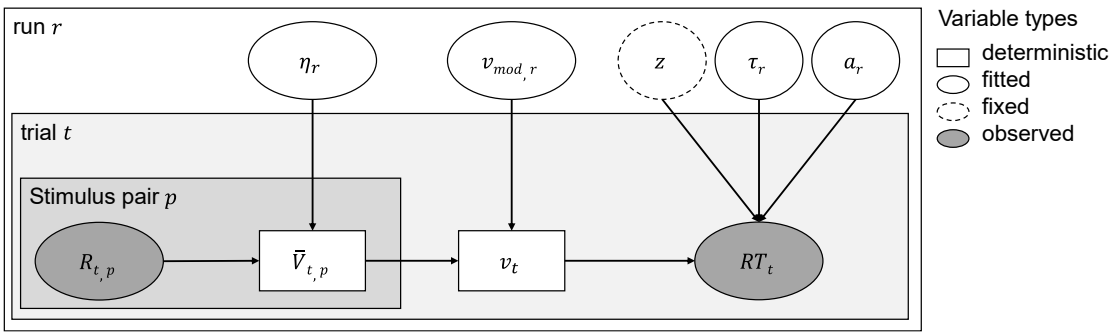
With this algorithm, the probability density for hitting the upper boundary (left choice) or lower boundary (right choice) at any given time point t could be calculated:

$$f(t|v,a,z) = \frac{\pi}{a^2} \exp\left(-vaz - \frac{v^2 t}{2}\right) \sum_{k=1}^{\infty} k \exp\left(-\frac{k^2\pi^2 t}{2a^2}\right) \sin(k\pi z) \tag{6}$$

This returns the probability of choosing x given the observed reaction time and the expected reward. For further details, please refer to the paper by Pedersen et al. (2017).

Figure 2

Schematic of the reinforcement learning drift diffusion model



Note. Deterministic variables are represented by rectangular nodes, fitted parameters by round nodes with solid lines, fixed parameters by round nodes with dashed lines, and observed data by

grey round nodes. The parameters are estimated for each run separately. Learning rates η modulate the trial-wise values $\bar{V}$ which are updated depending on the feedback. The difference between the value of the left and right stimulus pair is multiplied with the v_{mod} parameter to estimate the trial-wise drift rate v. Parameters are estimated for each run r. a = decision boundary; τ = nondecision time; z = starting point (fixed parameter); R = feedback; t = trial; p = audio-visual or tactile-visual stimulus pair; v = drift rate; RT = reaction time; r = run.

In our approach, the learning rate η was a free parameter whereas the initial values for all stimulus combinations were fixed to 0.5 (uniform unbiased initial values). The DDM included three additional free parameters: the non-decision time τ, the boundary separation a, and the drift weight v_{mod}. The starting point z in the DDM was fixed to $0.5 * a$, indicating no bias towards any of the decisions. The boundary separation indicates the amount of information required until a decision criterion is reached. The non-decision time represents the amount of time that is irrelevant for the decision-making process, such as the time required for response execution. The drift rate v indicates the rate of accumulation of information. In the approach used in this study, trial-wise values were obtained for the RPE, values, and drift rate, whereas the learning rate, boundary separation, and non-decision time had one estimated value per run.

Prior to applying the model to the dataset, we conducted simulations to verify that all model parameters could be recovered successfully (see Supplementary Material A.5). Then, the validated model was used for parameter estimation on the collected behavioural data and the estimated parameters were then used to identify corresponding neural processes (Farrell & Lewandowsky, 2018; R. C. Wilson & Collins, 2019). This approach ensured that our models accurately captured the underlying cognitive processes of multisensory learning. The model fitting pipeline was set up in MATLAB using the Genetic Algorithm (ga) from the Global Optimization Toolbox, in accordance with the recommendations set forth by R. C. Wilson & Collins (2019).

2.5 Statistical analyses

2.5.1 Analyses of task performance and computational parameters

As detailed in the supplementary materials (A.6), preprocessing of the behavioural data involved the exclusion of trials with omissions, and reaction times that were either too fast or too slow, resulting in a final set of 7,594 valid trials for analysis.

To evaluate learning outcomes and differences between modalities, mean reaction times for correct trials, mean accuracy scores (correct responses divided by the total number of trials not counting omissions) and mean absolute drift rate were calculated for each bin and modality, where one bin represented one-fourth of the task (11 trials). Reaction times, accuracies, learning rate, non-decision time, boundary separation, and the absolute drift rate were then analysed using linear mixed models (library 'lmerTest', Kuznetsova et al., 2017) with subject as a random effect and modality and age as fixed effects. The model for the reaction times, accuracies, and absolute drift rate also included bin as fixed effect. Post-hoc analyses for significant main effects with more than one level were computed (library 'emmeans'; Lenth, 2024) and p-values were adjusted for multiple comparisons by the Tukey method. Analyses were conducted using the R Statistical language (version 4.4.1; R Core Team, 2024).

2.5.2 Whole-brain analysis

For the first-level analysis, we built a general linear model (GLM) using the individual onsets of two regressors of interest (stimulus, feedback) and the two parametric modulators

for value and RPE. The value modulator was added on the stimulus regressor and included the normalised value for the chosen stimulus combination. The RPE modulator was added on the feedback regressor and included the trial-wise RPE values. We added 4 additional regressors of no interest: 1) stimulus onset for RT outlier trials, 2) feedback onset for RT outlier trials, 3) stimulus onset for omission trials, and 4) feedback onset for omission trials to the GLM. The GLM also included six realignment parameters and one additional regressor of no interest for each flagged scan, whenever available. This was done to account for movement effects (scrubbing). The GLM included all available AV and TV runs for each subject and was convolved with the canonical hemodynamic response function implemented in SPM12.

Whole-brain second-level one-sample t-tests investigated the activation during stimulus processing, feedback processing, value processing, and RPE processing. Additionally, age was added as an effect of interest to all contrasts. For the contrasts investigating stimulus processing, the learning rate was added as an additional effect of interest (if more than one run was included, the mean learning rate from all included runs was used). Further, to investigate main and interaction effects between age and modality (AV and TV), a mixed-effects ANCOVA was conducted. The design matrix included the scans for each condition and participant. Age, condition, and their interaction were included as regressors of interest, where age is a continuous and condition a factorial variable. The model accounted for variability within subjects by including one regressor per subject. These analyses additionally included the covariates sex and handedness to control for potential confounds. The effect of age, learning rate, and the described contrasts were used to investigate general brain activity during multisensory learning and how the activation differed between modalities (AV and TV) and in relation to age. Statistical parametric maps were thresholded at a voxel-wise $p<.001$ (uncorrected), with clusters corrected for multiple comparisons using a family-wise error (FWE) rate of $p<.05$, unless stated otherwise.

2.5.3 Region of interest analyses

For region of interest (ROI) analyses, the same first-level outputs were used as for the whole-brain analyses. Masks for multisensory regions, value processing and RPE processing regions were created using neurosynth (www.neurosynth.org). We downloaded the uniformity test activation mask for each topic and then processed each mask to extract all the subclusters as individual masks (a detailed description of the processing can be found in the Supplementary Material A.7). With the code provided by MarsBaR (https://marsbar-toolbox.github.io/faq.html), we extracted the beta values corrected for the mean signal in each ROI (Pernet, 2014). We extracted the mean beta values over all subclusters and for each subcluster separately from each mask. The “multisensory network” mask was used with the AV stimulus and TV stimulus contrasts, the “value network” mask with the AV value and TV value contrasts, and the “reward prediction error network” mask with the AV RPE and TV RPE contrasts. In a next step, these beta values were then analysed using linear mixed models in R to investigate differences between modalities and the effect of age. First, we looked at the effects of modality and age and their interaction on the whole mask (beta value ~ modality × age). Then we applied the same linear mixed model on each subcluster. In the multisensory network ROI, the learning rate was included as an additional predictor. To correct for multiple comparisons, all the p-values from each linear mixed model were extracted and corrected using the False Discovery Rate (FDR) procedure (Benjamini & Hochberg, 1995). The FDR correction was applied with a significance level of $\alpha = 0.05$. This allowed us to control for the proportion of false positives among the test results.

3 Results

3.1 Behavioural results

General Task Performance: the mean reaction times for AV and TV runs were 2.26s (*SD* = 0.37s) and 2.49s (*SD* = 0.35s), respectively. The median accuracy for AV and TV runs were 65.91% (*IQR* = 24.70%) and 65.12% (*IQR* = 23.58%), respectively. An overview including omissions and outliers can be found in the supplementary material B.2.

Linear mixed models were computed to examine the effect of sensory modality (AV vs TV), bin (1, 2, 3, 4) and age on reaction times and accuracy.

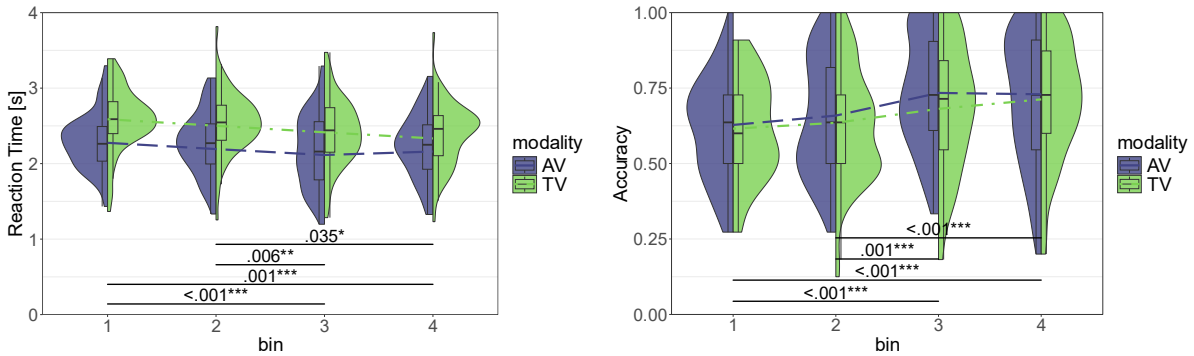
Reaction Time: The LMM for reaction times showed no significant main effect of modality (p = .814), but a significant main effect of bin (F(3,464) = 8.83, p < .001), and age (F(1,65) = 10.66, p = .002), and a significant interaction between modality and age (F(1,464) = 5.59, p = .018). Adding the interactions between modality and bin, age and bin, and the three-fold interaction between modality, age, and bin did not significantly improve model fit (*p*s > .0854). Post-hoc pairwise comparisons between the bins can be seen in Table B2.2 in the Supplementary Material. The results show significantly faster RTs in the third and last bin compared to the first and second bin. RTs were also faster the older the children, but this effect was weaker in TV compared to AV runs (see **Figure 3A** and **C**).

Accuracy: For the accuracy, the LMM showed significant main effects of modality (F(1,464) = 6.70, p = .010), bin (F(3,464) = 13.85, p < .001), and age (F(1,65) = 10.51, p = .002), and a significant interaction between modality and age (F(1,464) = 8.79, p = .003). Adding the interactions between modality and bin, age and bin, and the three-fold interaction between modality, age, and bin were did not significantly improve model fit (*p*s > .163). Post-hoc pairwise comparisons between the bins can be seen in Table B2.2 in the Supplementary Material. The results show significantly higher accuracies in the third and last bin compared to the first and second bin, and with increasing age of the children, although this effect is less strong in TV runs compared to AV runs (see **Figure 3B** and **D**).

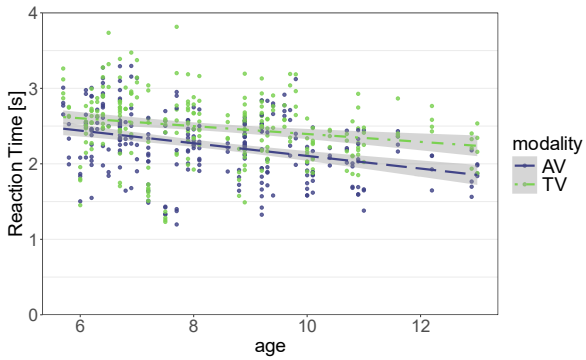
Figure 3

Reaction time and accuracy per modality and task bin

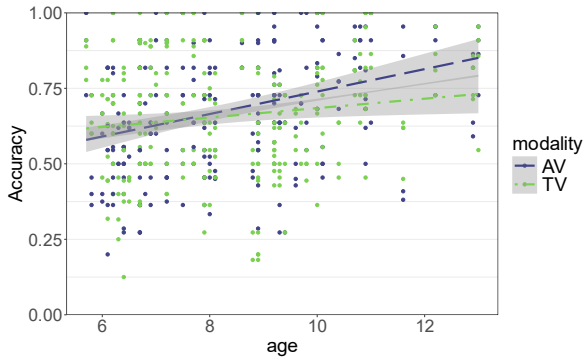
A Reaction time per modality and bin **B Accuracy per modality and bin**



C Interaction effect of modality and age on reaction time



D Interaction effect of modality and age on accuracy



Note. **A** shows the reaction times per task and bin of task on the x-axis **B** shows the accuracy per task and bin on the x-axis. The dashed and dotted lines connect the mean values for reaction times and accuracies per bin for AV and TV runs, respectively. **C** shows the interaction effect between age and modality on the reaction times. **D** shows the interaction effect between age and modality on the accuracies. AV = audio-visual (in blue), TV = tactile-visual (in green).

3.2 Modelling results

Linear mixed models were conducted to examine the effect of modality (AV vs. TV) and age on the learning rate, the non-decision time, and the boundary separation. For the drift rate, we fitted the same model as for the reaction time and accuracy, with modality (AV vs. TV), bin (1, 2, 3, 4), and age as predictors.

Learning rate: Results showed neither a significant main effect of modality nor of age on the learning rate (*p*s > .875).

Non-decision time: There was a significant main effect of modality on the non-decision time (F(1,66) = 8.67, p = .004) indicating significantly longer non-decision times and thus prolonged multisensory processing in TV compared to AV runs (see **Figure 4A**), but no significant effect of age (p = .190). Adding the interaction between modality and age to the model, did not improve model fit ($\chi^2(1)$ = 1.33, p = .250), indicating no significant interaction effect.

Boundary separation: There was a significant main effect of modality on the boundary separation (F(1,66) = 13.02, p < .001), indicating that participants require more evidence before reaching a decision in TV runs compared to AV runs (see **Figure 4B**). In contrast, the main effect of age was not significant (p = .125). Adding the interaction between modality and age to the model, did not improve model fit ($\chi^2(1)$ = 0.554, p = .457), suggesting no significant interaction effect.

Drift rate: Results on the drift rate show a significant main effect of modality (F(1, 461) = 9.44, p = .002) and age (F(1,65) = 12.30, p < .001), but not bin (p = .610). There was also a significant interaction between modality and age (F(1,461) = 15.12, p < .001) and between bin and age (F(3,461) = 8.70, p < .001). This indicates higher absolute drift rates and thus more efficient decision making in older children, but this age effect on decision efficiency is stronger in AV compared to TV runs and in later compared to earlier bins (see **Figure 4C and D**). Post-hoc pairwise comparisons between the bins can be seen in **Table 2**.

Table 2

Post-hoc tests of the effect of bin on drift rate

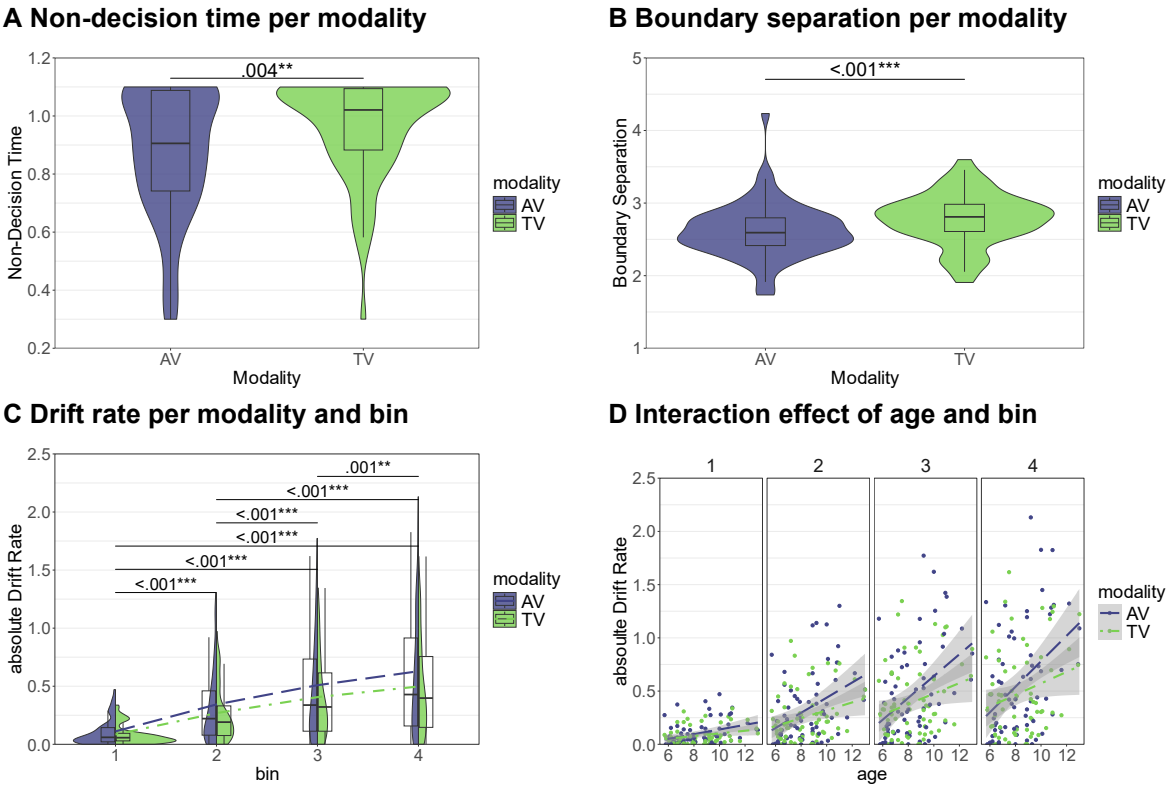
Model	contrast	estimate	SE	*df*	*t*	*p*
Drift rate ~ bin	**1st - 2nd**	**-0.20**	**0.03**	**461.00**	**-7.38**	**< .001*** **

1st - 3rd	**-0.34**	**0.03**	**461.00**	**-12.77**	**< .001*** **
1st - 4th	**-0.44**	**0.03**	**461.00**	**-16.48**	**< .001*** **
2nd - 3rd	**-0.14**	**0.03**	**461.00**	**-5.38**	**< .001*** **
2nd - 4th	**-0.24**	**0.03**	**461.00**	**-9.10**	**< .001*** **
3rd - 4th	**-0.10**	**0.03**	**461.00**	**-3.71**	**.001** **

Note. Post-hoc pairwise comparisons for the effect of the specified factor on the dependent variable. Estimates represent the mean difference between conditions. SE = standard error; df = degrees of freedom; p values are Bonferroni-corrected for multiple comparisons.

Figure 4

Modality and age effects on non-decision time, boundary separation, and drift rate



Note. **A** depicts the non-decision time and **B** the boundary separation for AV and TV runs. **C** shows the mean absolute drift rate per modality and bin. The dashed and dotted lines connect the mean values per bin for AV and TV runs, respectively. **D** plots the correlation between mean absolute drift rates and age per bin and modality. AV = audio-visual (in blue), TV = tactile-visual (in green).

3.3 Whole-brain fMRI results

To examine the effects of modality and age on brain activation, we analysed the BOLD signal during both stimulus and feedback processing using a whole-brain approach. Additionally, we applied parametric modulations during these phases to assess how factors such as values, and RPEs influence the BOLD signal.

3.3.1 Effects of age on stimulus and feedback processing

During AV and TV stimulus processing, age correlated positively with activation in an extended visual and parietal cluster, frontal regions, temporal regions, and the supplementary motor cortex. Brain activation in the bilateral middle orbital gyrus negatively correlated with age during AV stimulus processing. During TV stimulus processing, age correlated negatively with brain activation in temporal, hippocampal, frontal, and central regions (see **Figure 5A**). The main effect of age showed that the fusiform gyrus was more active with increasing age of the

children for both AV and TV runs and the calcarine gyrus less active the older the children were (see **Figure 5B**). The interaction effect indicates that the bilateral auditory and visual regions showed higher activation in older children in AV runs but not in TV runs. There were no significant clusters showing higher activation in older children in TV runs but not in AV runs (see **Figure 5C**). The cluster table and general brain activation during stimulus processing can be found in the Supplementary Material B.3.

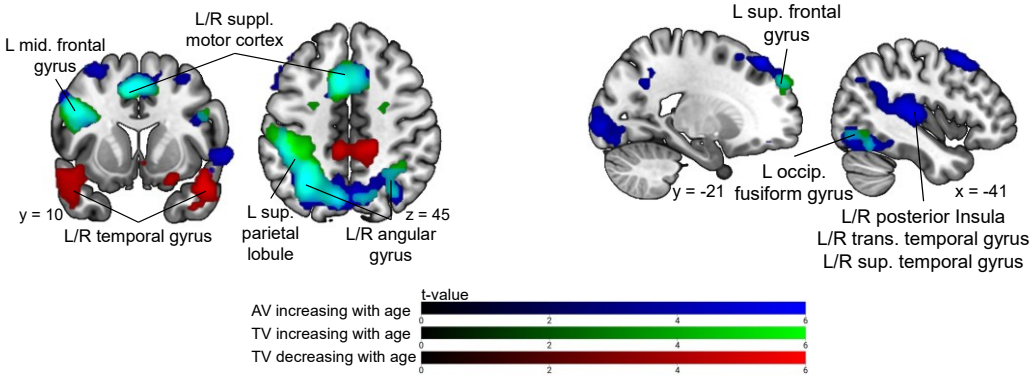
In AV runs during feedback processing, age positively correlated with brain activation increases in temporal, visual, and frontal regions. In TV runs during feedback processing, age correlated positively with the brain activation in visual and frontal regions (see **Figure 5A**). The main effect of age revealed that a cluster in the cerebellum showed higher activation with age in both AV and TV runs. However, it is important to note that full cerebellar coverage was not available for all participants. There were no significant clusters showing decreased activation with age in both AV and TV runs (see **Figure 5B**). The interaction effect between age and modality revealed that auditory regions showed higher activation with age in AV runs. There were no significant clusters showing higher activation in older children during TV runs, but not AV runs (see **Figure 5C**; the Table B4 in the Supplementary Material shows all the clusters). No significant negative correlation was observed during AV and TV feedback processing. The general brain activation during feedback processing can be found in the Supplementary Material B.4.

Figure 5

Age-effects on brain activation during stimulus and feedback processing

A Brain activation correlated with age for AV and TV runs

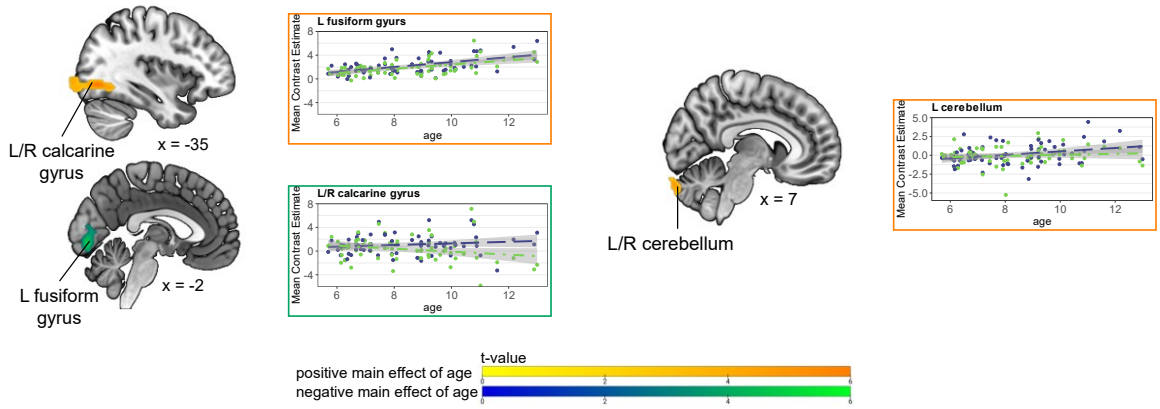
Stimulus processing **Feedback processing**



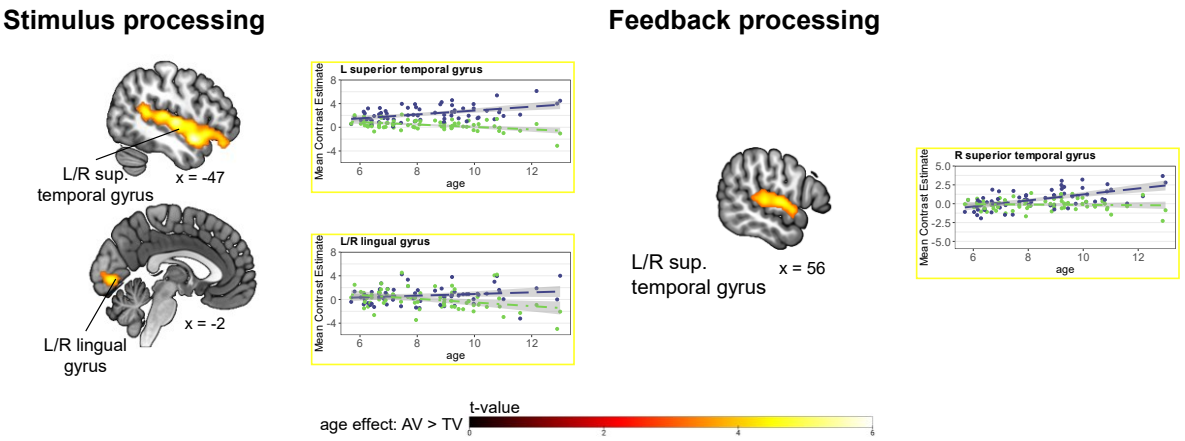
B Main effect of age for stimulus and feedback processing

Stimulus processing **Feedback processing**

Main effect of age Main effect of age



C Interaction effect of age and modality for stimulus and feedback processing



Note. **A** In dark blue, brain regions that show higher activation in older children during stimulus (left column) and feedback processing (right column) in AV runs. In green, brain regions showing higher activation in older children during stimulus and feedback processing in TV runs. In red, brain regions that are negatively correlated with age during stimulus processing in AV runs. Overlap between AV and TV in light blue. **B** positive (in yellow-orange) and negative (blue-green) main effect of age and interaction effects of age and modality (black-yellow) during stimulus (left column) and feedback processing (right column). And **C** the interaction effect of age and modality for stimulus (left) and feedback (right) processing showing higher brain activation with age in AV runs. The correlation plots show the mean estimate contrast values for each subject and modality in the corresponding clusters. AV in blue, TV in green. Cluster defining threshold p<.001 (unc.), p_{FWEc} = .05, family-error wise corrected. AV = audio-visual, TV = tactile-visual, L = left, R = right.

3.3.2 Modulation of the BOLD signal by model parameters

Significant modulation of the activation was observed for the parametric modulation of the BOLD signal with reward prediction errors and value processing. Additionally, the BOLD signal correlated significantly with the learning rate but not with drift rate. These findings are further explored in Sections A–C below.

A) Reward prediction error processing

Whole-brain analyses showed no significant differences in RPE processing between AV and TV runs. Therefore, for subsequent analyses AV and TV runs were pooled. RPE positively modulated the BOLD activity during feedback processing in the ventral striatum, the precuneus, and in medial frontal regions, and negatively in an extended cluster around the anterior insula, the supplementary motor cortex and superior frontal cortex, right temporal regions, and the right fusiform gyrus (see **Figure 6A** and **Table 3**). We did not find a significant effect of age on RPE processing in the whole-brain analyses.

Table 3

Significant clusters showing a positive and negative modulation by RPEs during feedback processing

Contrast	Brain Area	MNI Coordinates			Cluster size	T-value	Peak-level p_{unc}	Cluster-level p_{FWE}
		x	y	z				
AV + TV	R Amygdala	13	0	-15	1185	8.46	<.001	<.001
	R Hippocampus	16	-39	15	984	5.78	<.001	<.001
	L Middle Frontal Gyrus	-26	21	48	151	4.96	<.001	.029
	R Insula Lobe	34	21	6	1604	-8.36	<.001	<.001
	R middle cingulate cortex	7	24	36	1405	-7.33	<.001	<.001

L inferior frontal gyrus (p. Orbitalis)	-29	21	-9	631	-7.28	<.001	<.001
R Fusiform Gyrus	28	-66	-9	371	-6.16	<.001	<.001
R Middle Temporal Gyrus	52	-33	-3	522	-5.95	<.001	<.001
L Superior Temporal Gyrus	-62	-48	27	147	-4.74	<.001	.032

Note. Only the main peak is shown in this table. Regions were automatically labeled using the AnatomyToolbox atlas. x, y, and z =Montreal Neurological Institute (MNI) coordinates in the left-right, anterior-posterior, and inferior-superior dimensions, respectively. Cluster defining threshold p<.001 (unc.), p_{FWEc} = .05, family-error wise corrected. Abbreviations: L = left, R = right.

B) Value processing

Whole-brain analyses showed no significant differences in value processing between AV and TV runs. Therefore, for subsequent analyses AV and TV runs were pooled. Brain activity in the left pre- and postcentral gyrus during stimulus processing was positively modulated by the value signal (see **Figure 6B** and **Table 4**). There were no significant differences between AV and TV runs and no effect of age. We did not find a significant effect of age on value processing.

Table 4

Significant clusters showing a positive modulation by values

Contrast	Brain Area	MNI Coordinates			Cluster size	T-value	Peak-level p_{unc}	Cluster-level p_{FWE}
		x	y	z				
AV + TV	L middle cingulate cortex	-11	-15	54	218	5.06	<.001	.002
	L Precentral Gyrus	-35	-24	54		4.31	<.001	
	L Precentral Gyrus	-38	-21	66		4.00	<.001	

Note. Regions were automatically labeled using the AnatomyToolbox atlas. x, y, and z =Montreal Neurological Institute (MNI) coordinates in the left-right, anterior-posterior, and inferior-superior dimensions, respectively. Cluster defining threshold p<.001 (unc.), p_{FWEc} = .05, family-error wise corrected. Abbreviations: L = left, R = right.

C) Neural correlates of the learning rate

The mean learning rate, calculated across all runs for each subject, positively correlated with mean brain activation during stimulus processing, averaged over all runs, in the medial frontal cortex (see **Figure 6C** and **Table 5**). No correlations between mean learning rate and stimulus activation were observed when looking at AV and TV runs individually.

Table 5

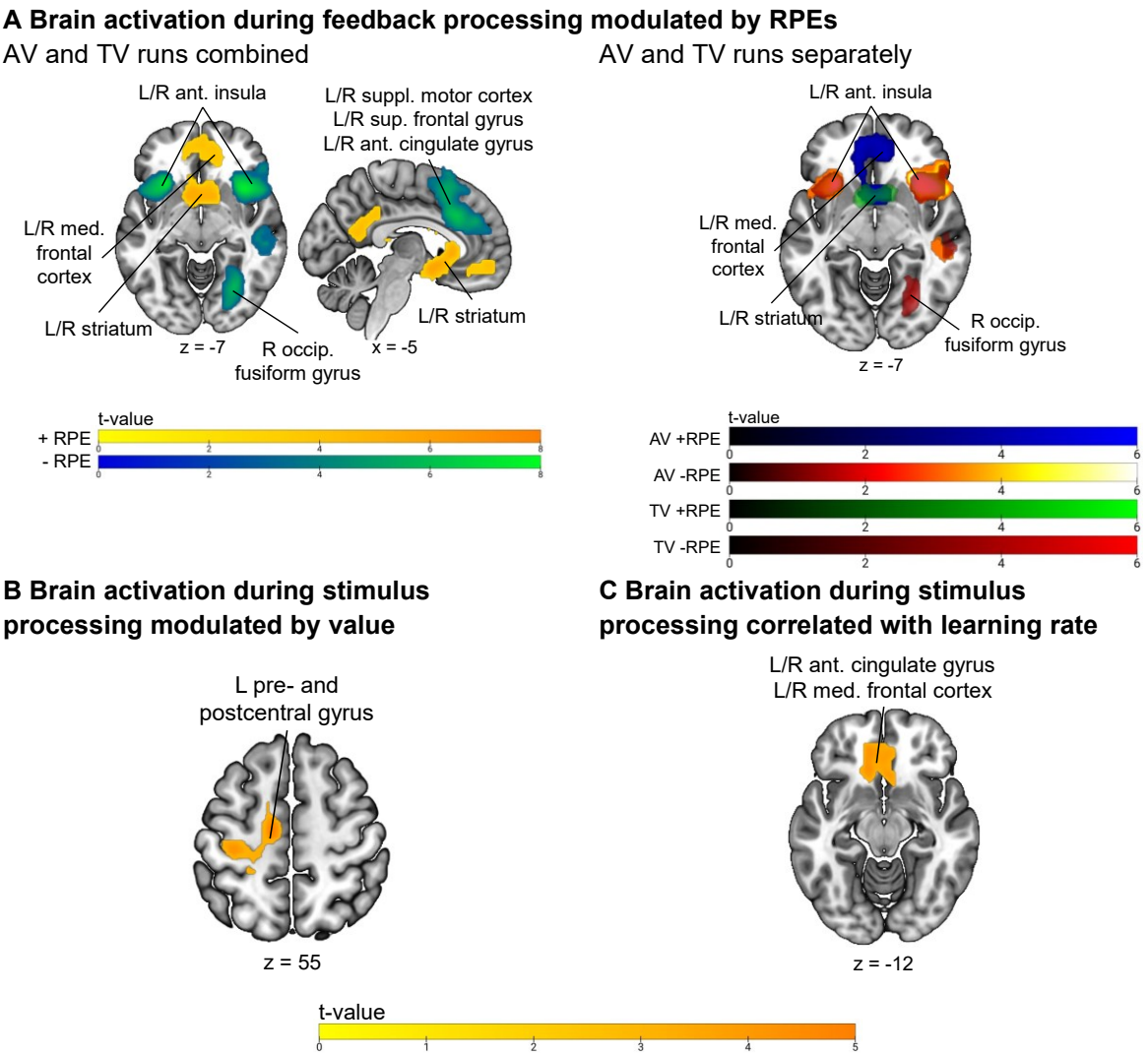
Significant clusters showing a positive correlation with learning rate

Contrast	Brain Area	MNI Coordinates			Cluster size	T-value	Peak-level p_{unc}	Cluster-level p_{FWE}
		x	y	z				
AV + TV	R Mid Orbital Gyrus	1	24	-9	228	4.44	<.001	.002

Note. Only the main peak is shown in this table. Regions were automatically labelled using the AnatomyToolbox atlas. x, y, and z = Montreal Neurological Institute (MNI) coordinates in the left-right, anterior-posterior, and inferior-superior dimensions, respectively. Cluster defining threshold p<.001 (unc.), p_{FWEc} = .05, family-error wise corrected. Abbreviations: L = left, R = right.

Figure 6

Effects of reward prediction errors, values, and learning rate on brain activation



Note. **A** shows the brain activation during feedback processing positively and negatively modulated by RPEs for AV and TV runs combined (left column) and separately for both modalities (right column; results reported in Supplementary Material B.5). In yellow-orange, brain regions that are positively modulated by the RPE. In blue-green, brain regions that are negatively modulated by the RPE. In dark blue the positive modulation during AV runs, in black-yellow the negative modulation during AV runs. In green the positive modulation during TV runs and in red the negative modulation during TV runs. **B** shows the brain activation related to values during stimulus processing for AV and TV runs combined. **C** shows the brain activation during stimulus processing (AV and TV runs combined) correlated with the learning rate. In yellow-orange, brain regions that are positively correlated with/modulated by the respective parameter. L = left, R = right, RPE = reward prediction error, AV = audio-visual, TV = tactile-visual. Cluster defining threshold p<.001 (unc.), p_{FWEc} = .05, family-error wise corrected.

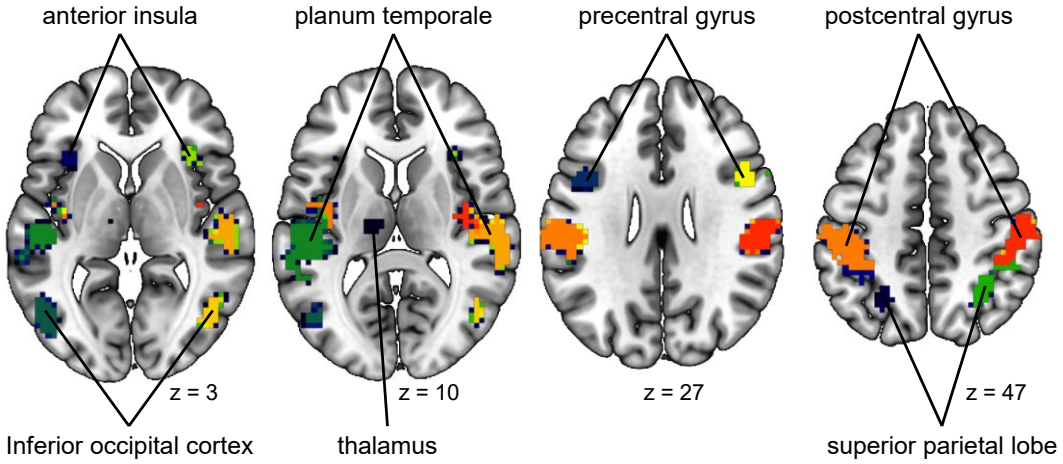
3.4 ROI analyses

To examine the effects of age and modality more closely, we conducted ROI analyses in a multisensory, a RPE, and a value mask as well as for each region included in these three masks separately. The masks and included regions used for these analyses are shown in **Figure 7**. ROI analyses of value processing are presented in Supplementary Material B.8, as whole-brain analyses revealed no significant activation within the value mask.

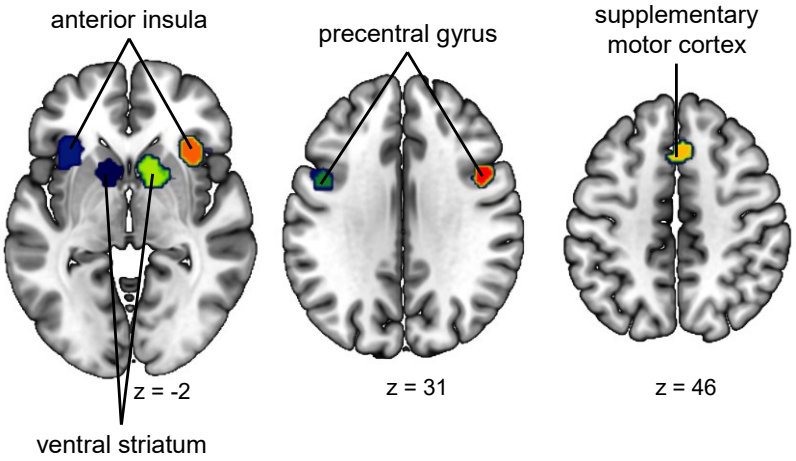
Figure 7

Depiction of meta-analytical masks used for ROI analyses

A Multisensory network



B Reward prediction error network



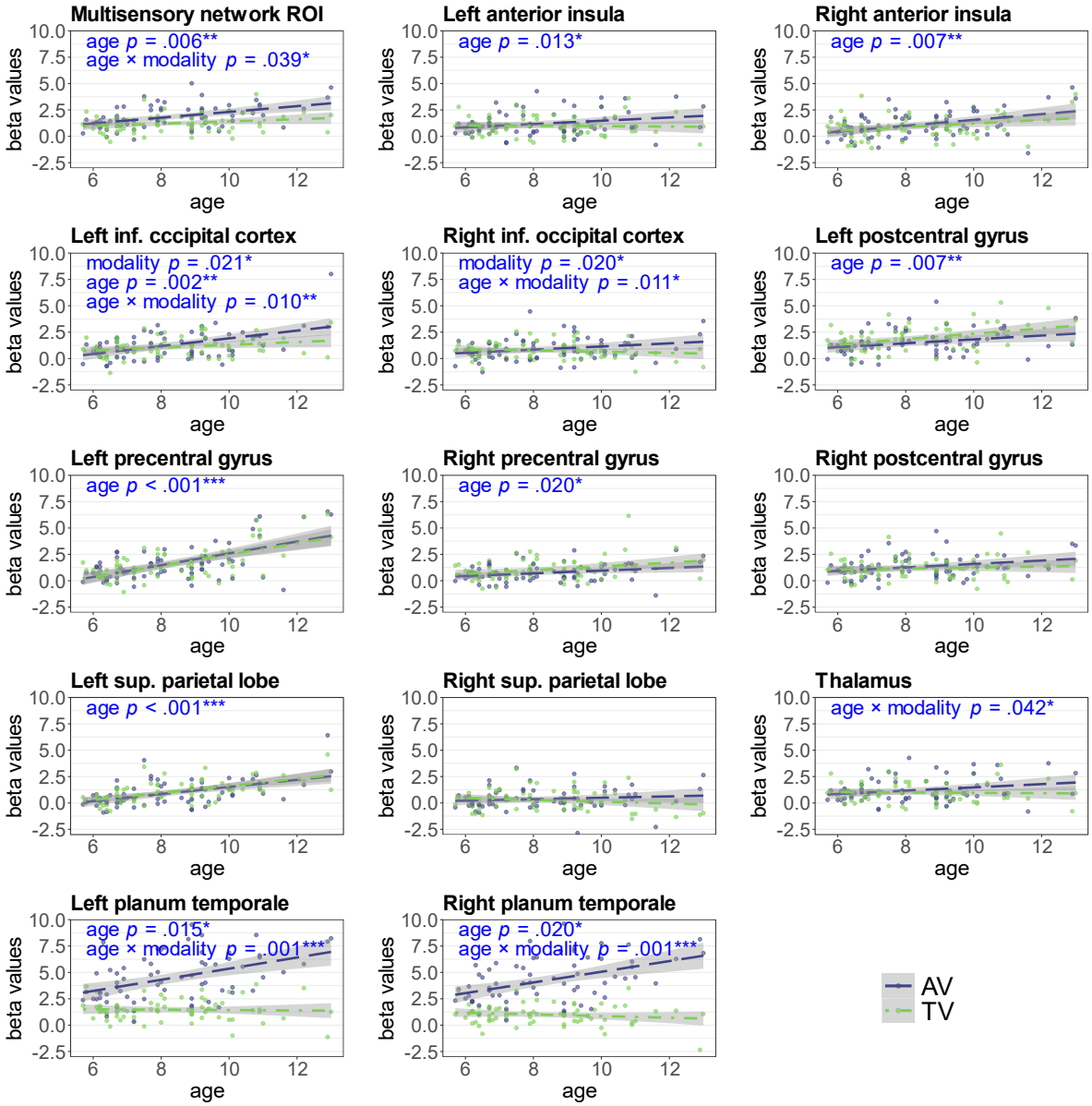
Note. Panel **A** shows multisensory clusters and **B** the RPE clusters derived from NeuroSynth.

3.4.1 Brain activation during stimulus presentation

In the "multisensory network ROI" which comprises multiple distinct multisensory clusters, brain activation was higher in older children, with the interaction indicating a weaker effect of age in TV runs compared to AV runs (see **Figure 8**). All separate clusters except the right postcentral gyrus, the right superior parietal lobule and the left thalamus show higher activation with increasing age of the children. This effect is less strong in TV runs compared to AV runs in the left and right inferior occipital cortex and the left and right planum temporale. In the left thalamus there is no main effect of age, but in TV runs we see lower activation in older children compared to younger children. In the left and right inferior occipital cortex, we also observed a main effect of modality (see **Figure 8**). The learning rate did not influence brain activation during stimulus processing (the table with the effects can be found in Supplementary Material B.6).

Figure 8

Beta values in the multisensory network ROI and distinct clusters



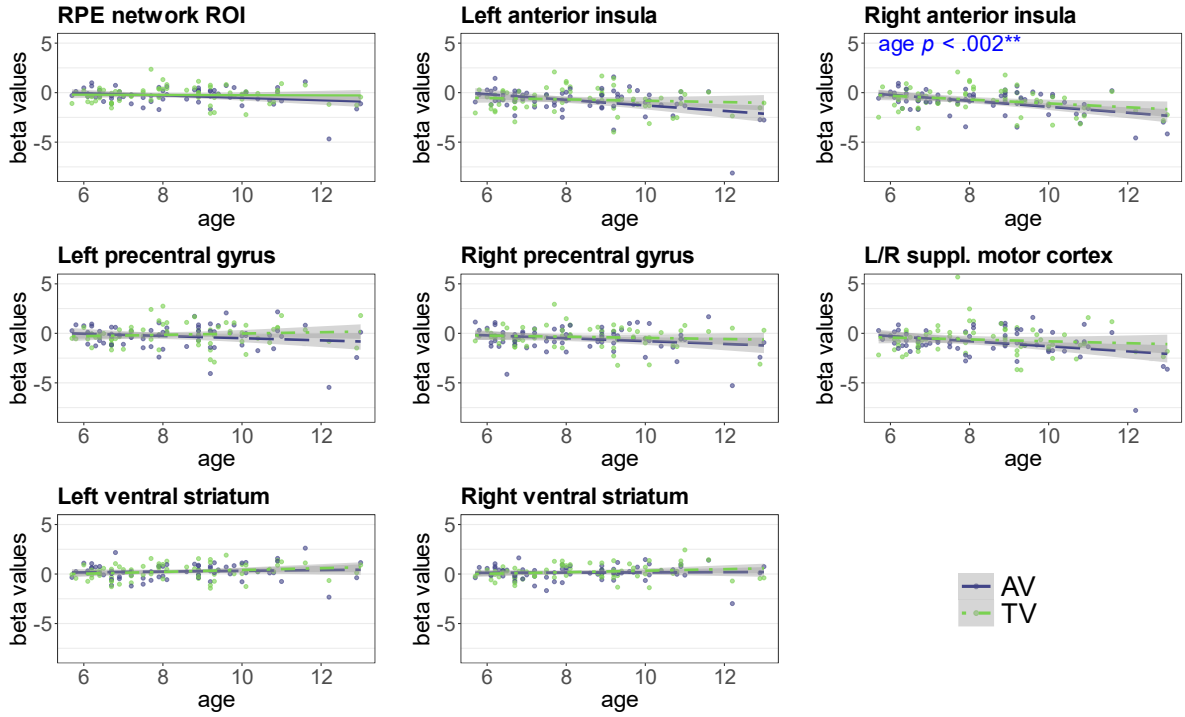
Note. This figure shows the mean beta values for AV and TV stimulus processing in the whole multisensory network ROI and in each separate cluster, per subject and modality. AV in blue, TV in green. *P*-values are FDR corrected for multiple comparisons.

3.4.2 Brain activation during reward prediction error processing

For the whole RPE Network ROI, mean activation significantly differed from zero, but otherwise there are no significant effects. Importantly, the right anterior insula shows higher brain activation in older children to lower RPEs (see **Figure 9**; the table with model summaries can be found in Supplementary Material B.7).

Figure 9

Beta values in the reward prediction error network ROI and distinct clusters



Note. This figure shows the mean beta values for the RPE contrasts in the RPE Network ROI and each separate cluster per subject and modality. AV in blue, TV in green. *P*-values are FDR corrected for multiple comparisons.

4 Discussion

This study provides insights into the behavioural and neural processing underlying multisensory learning in children, focusing on audio-visual and tactile-visual associations. Three key findings emerged: First, tactile-visual associations were harder to learn than audio-visual associations as shown by performance data across ages. Second, despite considerable individual variability in learning performance, a clear developmental pattern emerged: Older-aged children demonstrated faster processing speed and greater decision-making efficiency, and this enhanced behavioural performance corresponded to higher activation in brain regions for sensory processing, cognitive control, MSI, and memory retrieval. This effect was more pronounced during AV learning as evident by the significant age-effects during AV runs on neural processing which were not present during TV runs, aligning with known neural maturation processes that support multisensory learning as children develop. Third, while the core processes of reward-based learning and RPE processing, which were modality-independent, remained relatively stable across middle childhood, the older the children were, the higher their brain activity for more negative RPE, suggesting an increased sensitivity to negative feedback. This may indicate a more refined approach to evaluating rewards and punishments with development and suggests a developmental shift in how children process and utilize feedback information to guide their learning.

These findings reveal modality- and age-related alterations in behavioural and neural processes during multisensory learning, highlighting stable core learning processes and developmental shifts in neural engagement. They further emphasize the importance of accounting for age and modality in MSI research.

4.1 Modality-specific processing and learning across middle childhood

Behavioural analyses revealed significant learning during the multisensory task across both modalities. This improvement was shown by faster reaction times and higher accuracies in the second half of each experimental run. We observed distinct differences in performance between the two sensory modalities. Tactile-visual runs consistently yielded lower accuracies compared to audio-visual runs, indicating that associations were more challenging for children to learn. This modality-specific difference in learning difficulty became more pronounced with increasing age. Computational modelling provided deeper insights into the learning and decision-making in multisensory learning, clarifying modality-specific and age-related changes. First, we discuss how the sensory modality influence learning processes and second how age affects learning across modalities.

The learning rate, which reflects the influence of RPEs on value updating (Zhang et al., 2020), showed similar speed of learning regardless of sensory modality. However, other computational parameters differed significantly between modalities. Specifically, non-decision time was longer, decision boundaries were wider, and drift rates were lower in TV compared to AV learning. These findings align with slower reaction times for tactile-visual association learning, suggesting that children required more information accumulation and showed a slower accumulation process for decision-making in TV runs (Friedrich & Beste, 2019; Ratcliff & McKoon, 2008).

Consistent with the response time analyses, computational modelling revealed that processing speed improved with increasing age, driven by increased drift rates. Older children had higher drift rates in both audio-visual and tactile-visual modalities, with a more pronounced increase in audio-visual learning, reflecting faster and more efficient information accumulation during decision processes (Ratcliff & Rouder, 1998). Non-decision time and decision

boundaries remained stable across ages, indicating that the additional processing demands of tactile encoding were consistent across age. Additionally, thresholds for accumulating information did not vary with age. In summary, while learning rates were consistent across modalities and ages, modality influenced non-decision time, boundary separation, and drift rate, with tactile-visual pairings requiring more extensive processing. In contrast, age enhanced processing speed and decision efficiency across modalities without altering the task's modality-specific demands. This suggests that maturation enhances children's ability to associate relevant sensory information, improving speed and accuracy in multisensory learning tasks (Barutchu et al., 2020). To investigate the neural underpinnings of these age-related behavioural differences, we analysed brain activation patterns during stimulus and feedback processing at the whole-brain level and in specific ROIs.

4.2 Age-related recruitment of higher-order brain regions during multisensory learning

Whole-brain analyses showed a negative correlation between age and lingual gyrus activation during stimulus processing, which was driven by a reduction of activation during TV association learning with age, suggesting a reliance on basic visual features for stimulus discrimination in younger children (Bogousslavsky et al., 1987; Murray et al., 2016; Palejwala et al., 2021). While younger children also recruit higher-order regions for MSI, they may do so less efficiently than older children, highlighting ongoing MSI network maturation. In contrast, brain activation in the inferior occipital gyrus during stimulus processing in both AV and TV runs correlated positively with age. This is a region involved in matching visual stimuli by perceptual and semantic features (e.g., matching 'piano' to 'harp' instead of 'oven'; Ludersdorfer et al., 2019). This suggests that older children may engage in more detailed visual analyses to integrate visual, auditory, and tactile information.

Brain activation during stimulus processing increased with age in the occipital cortex, precuneus, the supplementary motor area (SMA), middle frontal gyrus, and parietal cortex. This increased neural activity reflects the increased engagement of neural resources in older children to successfully perform the demanding learning task. Enhanced intraparietal sulcus activity is linked to cross-modal integration (Sours et al., 2017), mnemonic processes during both multimodal encoding and retrieval (Tibon et al., 2019), and attention (Boeken & Markett, 2023), indicating a more focused attention on relevant visual stimuli supporting encoding and learning of multisensory associations. The angular gyrus, a cross-modal hub (Hirst et al., 2021; Seghier, 2013; Tibon et al., 2019), integrates information from multiple sensory modalities, which is crucial for forming multisensory associations. The superior parietal lobule, implicated in attention processes, supports the selection, and focus on relevant sensory information during learning (Alahmadi, 2021). Greater activity in the SMA, associated with motor planning (Chung et al., 2005) and working memory (Bahlmann et al., 2009; Chung et al., 2005; Cona & Semenza, 2017), suggests that older children may plan their responses to a greater extent which corresponds to the higher accuracies in older children. Additionally, the increased middle frontal gyrus activity, tied to working memory and cognitive control (Eichenbaum, 2017), implies a more efficient knowledge access, retrieval, and utilisation during multisensory learning. Modality-specific age-related increases were observed in the superior temporal lobe during AV stimulus processing involved in auditory perception and short-term auditory memory (Crinion et al., 2006; Joseph, 1990). This points towards a greater reliance of older children on auditory memory to consolidate and retrieve learned associations.

Interestingly, modality-specific age-related decreases in activation were observed primarily in the TV condition. Younger children exhibited higher activation in the anterior middle

temporal gyrus and hippocampus during TV processing, suggesting increased engagement in semantic processing (Rolls, 2021). This indicates that younger children may associate semantic meanings to tactile stimuli to enhance memory and retrieval.

The results of the whole-brain analyses were corroborated by ROI analyses, focusing on regions identified as relevant for multisensory processing and integration in a Neurosynth meta-analysis. Older children exhibited greater activation in the anterior insula, planum temporale, and inferior occipital cortex, particularly during AV runs. These results show that older children engage sensory processing, cognitive control, MSI, attention allocation and memory regions more extensively. This enhanced neural engagement likely reflects greater cognitive effort, improving performance in multisensory learning tasks.

4.3 Differential brain activation in AV and TV feedback processing with age

On a whole-brain level, brain activation during AV and TV feedback processing positively correlated with age in the superior frontal gyrus. This region, anatomically connected to the middle and inferior frontal gyri, supports working memory, episodic memory, and attention (Boisgueheneuc et al., 2006; Diveica et al., 2023; Li et al., 2013). Modality specific analyses revealed a positive correlation between brain activation in auditory regions and age during visual feedback processing in AV runs, suggesting older children utilize auditory memory traces to evaluate feedback and adjust behaviour to a greater extent. Similarly, increased left fusiform gyrus activation with age in AV and TV runs indicates an increased reliance on visual memory traces for the integration of multisensory information. The fusiform gyrus, crucial for visual attention (Chen et al., 2019; Ruiz-Rizzo et al., 2018), and letter and word form processing (Dehaene & Cohen, 2011; Thesen et al., 2012), facilitates the formation of audio-visual percepts, such as letter-speech sound associations (Karipidis et al., 2018, 2021; Pleisch et al., 2019). In contrast, age-related correlations during TV feedback processing were minimal, likely due to the increased difficulty of integrating tactile information, which limited age-related effects. This aligns with lower TV performance and suggests similar feedback processing and learning strategies across ages. The fMRI results further align with the drift rate findings, showing more pronounced age-related changes in the AV than TV learning.

4.4 Modality-independent value and reward prediction error processing

Our findings revealed a positive correlation between activation in the pre- and post-central gyri and value processing, consistent across both audio-visual and tactile-visual runs. This diverges from prior studies suggesting that the orbitofrontal gyrus, nucleus accumbens, insula, anterior cingulate cortex, and supplementary motor area are key for value processing (Liu et al., 2011; R. P. Wilson et al., 2018). The pre- and post-central gyri, involved in motor planning (Ariani et al., 2025; De Lange et al., 2005), may reflect increased readiness or preparatory activation for action, facilitating quicker behavioural responses. The lack of significant differences between AV and TV runs suggests that value processing in these regions is sensory-independent, as previously reported (Gao & Shinkareva, 2021). Similarly, RPE processing was modality-independent, with RPEs positively modulating brain activity in the ventral striatum, known for encoding RPEs, as well as the medial frontal cortex and hippocampus, linked to learning and memory (Cohen & Ranganath, 2005; Fraga-González et al., 2025; Garrison et al., 2013; McClure et al., 2003; O'Doherty et al., 2003; Pagnoni et al., 2002). RPEs negatively modulated brain activation in regions typically associated with negative RPE signalling, such as the cingulate cortex, anterior insula, and frontal areas (Fouragnan et al., 2018). ROI analyses further confirmed the modality independent RPE

processing. While no age-related effects on RPE processing were found in whole-brain analyses, ROI analyses revealed a negative correlation between age and brain activity in the right anterior insula. The findings of higher anterior insula activation with negative RPEs in older children are indeed consistent in previous findings in adolescents (Hauser et al., 2015; Yaple et al., 2020). This observation supports neurodevelopmental theories proposing that the anterior insula, as an immaturely connected cognitive-emotional hub in adolescence (Smith et al., 2014), biases decision-making towards affectively influenced choices, consistent with the concept of a dominant social-affective system in adolescent cognition (Crone & Dahl, 2012; Icenogle & Cauffman, 2021; Loureiro, 2020; Smith et al., 2014; Van Leijenhorst et al., 2010). For our MSL paradigm the findings thus suggests that older children exhibit heightened sensitivity to negative or unexpected feedback, processing it more extensively, thereby facilitating more effective learning adjustments and error correction. Finally, learning rate correlated with activation in the anterior cingulate cortex and the medial frontal cortex, regions involved in updating expectations and adjusting behaviour (Cohen & Ranganath, 2007; Debener et al., 2005; Hester et al., 2008; Matsumoto et al., 2007; O'Reilly, 2013; Rushworth, 2008; Wu et al., 2017). This aligns with the fact that higher learning rates are linked to faster and more substantial updating of beliefs (Zhang et al., 2020). In summary, as development progresses children exhibit enhanced MSI, relying on a broader network of brain regions, particularly during audio-visual learning. While value and RPE processing are modality-independent and stable across ages, older children demonstrate increased sensitivity to negative feedback, suggesting more efficient learning from reward and punishment.

4.5 Limitations

Our study offers novel insights into modality and age-related differences in neural activation during multisensory learning in children, but several limitations should be considered.

First, individual differences in learning abilities, attention and motivation may have influenced performance, particularly among younger children. Eye-tracking in future studies could help assess attention and engagement more directly.

Second, the adaptive nature of the task, ensuring comparable demands across participants, may have masked subtle age-related differences in cognitive and sensory processing. The task's inherent difficulty, especially for younger children, may have limited performance and contributed to variability in learning.

Third, although we designed the tactile vibration patterns to be distinct, they were generally more difficult to differentiate than the auditory stimuli, likely explaining the lower accuracy values during tactile-visual learning. Last, as a cross-sectional study, our research captures a snapshot of performance at a single developmental stage. Longitudinal studies would be beneficial to track the evolution of these cognitive processes in individual children and facilitate the understanding the developmental trajectories involved in multisensory learning.

4.6 Conclusion

Our findings show that typically developing children can learn new associations between abstract visual and auditory or tactile stimuli within an experimental session. Learning rates remained consistent across modalities and age groups, but modality-specific factors influenced decision-making. With increasing development, children exhibited more efficient multisensory processing and integration. This was accompanied by increased neural activation in regions linked to cognitive control, working memory, and MSI, indicating the

usage of increased neural resources for MSL the older the children. Older children exhibited increased activation in the right anterior insula in response to negative RPEs, suggesting a heightened sensitivity to negative feedback and a more nuanced learning process that integrates both positive and negative feedback for behavioural adaptation. In contrast, younger children in our study apparently relied more on basic learning mechanisms and seemed less able to flexibly adapt their behaviour based on negative feedback. Overall, older children use advanced cognitive strategies and neural resources in multisensory learning tasks, while younger children engage higher-order MSI regions less efficiently, reflecting ongoing dynamics and developmental maturation in multisensory learning. To conclude, our findings indicate that with increasing age children are more efficient at integrating information across different sensory modalities to aid learning of multisensory associations, which they may utilize more effectively to guide their behaviour. This ability is essential for advanced learning, problem-solving, and adaptation in dynamic environments.

CRediT authorship contribution statement

Nina Raduner: conceptualisation, data curation, formal analysis, investigation, methodology, project administration, visualization, writing – original draft; **Carmen Providoli:** conceptualisation, data curation, investigation, project administration, writing – review and editing; **Sarah Di Pietro:** conceptualisation, data curation, investigation, project administration, supervision, writing – review and editing; **Maya Schneebeli:** methodology, supervision, writing – review and editing; **Iliana I. Karipidis:** conceptualisation, supervision, writing – review and editing; **Ella Casimiro:** conceptualisation, methodology; **Saurabh Bedi:** conceptualisation, methodology, writing – review and editing; **Michael Von Rhein:** conceptualisation, funding acquisition, supervision, writing – review and editing; **Nora Maria Raschle:** conceptualisation, funding acquisition, methodology, supervision, writing – review and editing; **Christian C. Ruff:** conceptualisation, funding acquisition, methodology, resources, supervision, writing – review and editing; **Silvia Brem:** conceptualisation, funding acquisition, methodology, project administration, resources, supervision, writing – review and editing.

Declaration of competing interest

The authors report no competing financial interests or personal affiliations that may have affected the research presented in this paper.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT3.5 in order improve language and readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Acknowledgments

We are very grateful for all the families and children for participating in this study. Further, we thank K. Bremer, A. Bussmann, N. Ehrhardt, E. Hohl, K. Gelic, S. Gubler, L. Hämmerli, S. Ismail, R. Jocham, I.P. Johnson, A. Lütolf, X. Michael, M. Peter, S. Schädler, J. Schiesser, M.A. Steuer, P. Sting, L. Strickler, C. Volk, and R. Wombacher for their help in planning, recruitment, and data collection. We thank Y. Cao for his contribution to implementing the drift diffusion model, P. Stämpfli for his advice on the MRI protocol, G.S.P. Pamplona and A. Haugg for their help in the statistical analyses and R. Borbás, P. Dimanova and S. Scatolin for their valuable input.

Funding

This work was supported by the University Research Priority Program Adaptive Brain Circuits in Development and Learning (AdaBD) of the University of Zurich (project ChildBrainCircuits).

References

Alahmadi, A. A. S. (2021). Investigating the sub-regions of the superior parietal cortex using functional magnetic resonance imaging connectivity. *Insights into Imaging*, *12*(1). https://doi.org/10.1186/s13244-021-00993-9

Altarelli, I., Dehaene-Lambertz, G., & Bavelier, D. (2020). Individual differences in the acquisition of non-linguistic audio-visual associations in 5 year olds. *Developmental Science*, *23*(4). https://doi.org/10.1111/desc.12913

Ariani, G., Shahbazi, M., & Diedrichsen, J. (2025). Cortical areas for planning sequences before and during movement. *Journal of Neuroscience*, *45*(3).

Bahlmann, J., Schubotz, R. I., Mueller, J. L., Koester, D., & Friederici, A. D. (2009). Neural circuits of hierarchical visuo-spatial sequence processing. *Brain Research*, *1298*. https://doi.org/10.1016/j.brainres.2009.08.017

Barutchu, A., Crewther, S. G., Fifer, J., Shivdasani, M. N., Innes-Brown, H., Toohey, S., Danaher, J., & Paolini, A. G. (2011). The relationship between multisensory integration and IQ in children. *Developmental Psychology*, *47*(3), 877–885. https://doi.org/10.1037/a0021903

Barutchu, A., Fifer, J. M., Shivdasani, M. N., Crewther, S. G., & Paolini, A. G. (2020). The Interplay Between Multisensory Associative Learning and IQ in Children. *Child Development*, *91*(2), 620–637. https://doi.org/10.1111/cdev.13210

Barutchu, A., Toohey, S., Shivdasani, M. N., Fifer, J. M., Crewther, S. G., Grayden, D. B., & Paolini, A. G. (2019). Multisensory perception and attention in school-age children. *Journal of Experimental Child Psychology*, *180*, 141–155. https://doi.org/10.1016/j.jecp.2018.11.021

Batson, M. A., Beer, A. L., Seitz, A. R., & Watanabe, T. (2011). Spatial shifts of audio-visual interactions by perceptual learning are specific to the trained orientation and eye. *Seeing and Perceiving*, *24*(6). https://doi.org/10.1163/187847611X603738

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *57*(1). https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Birch, H. G., & Belmont, L. (1965). AUDITORY-VISUAL INTEGRATION, INTELLIGENCE AND READING ABILITY IN SCHOOL CHILDREN. *Perceptual and Motor Skills*, *20*, 295–305. https://doi.org/10.2466/pms.1965.20.1.295

Boeken, O. J., & Markett, S. (2023). Systems-level decoding reveals the cognitive and behavioral profile of the human intraparietal sulcus. *Frontiers in Neuroimaging*, *1*. https://doi.org/10.3389/fnimg.2022.1074674

Bogousslavsky, J., Miklossy, J., Deruaz, J. P., Assal, G., & Regli, F. (1987). Lingual and fusiform gyri in visual processing: A clinico-pathologic study of superior altitudinal hemianopia. *Journal of Neurology, Neurosurgery and Psychiatry*, *50*(5). https://doi.org/10.1136/jnnp.50.5.607

Boisgueheneuc, F. Du, Levy, R., Volle, E., Seassau, M., Duffau, H., Kinkingnehun, S., Samson, Y., Zhang, S., & Dubois, B. (2006). Functions of the left superior frontal gyrus in humans: A lesion study. *Brain*, *129*(12). https://doi.org/10.1093/brain/awl244

Butler, A. J., James, T. W., & James, K. H. (2011). Enhanced multisensory integration and motor reactivation after active motor learning of audiovisual associations. *Journal of Cognitive Neuroscience*, *23*(11). https://doi.org/10.1162/jocn_a_00015

Chen, L., Wassermann, D., Abrams, D. A., Kochalka, J., Gallardo-Diez, G., & Menon, V. (2019). The visual word form area (VWFA) is part of both language and attention circuitry. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-13634-z

Chien, S., Wiehler, A., Spezio, M., & Gläscher, J. (2016). Congruence of inherent and acquired values facilitates reward-based decision-making. *Journal of Neuroscience*, *36*(18). https://doi.org/10.1523/JNEUROSCI.3084-15.2016

Chung, G. H., Han, Y. M., Jeong, S. H., & Jack, C. R. (2005). Functional heterogeneity of the supplementary motor area. *American Journal of Neuroradiology*, *26*(7).

Cohen, M. X., & Ranganath, C. (2005). Behavioral and neural predictors of upcoming decisions. *Cognitive, Affective and Behavioral Neuroscience*, *5*(2). https://doi.org/10.3758/CABN.5.2.117

Cohen, M. X., & Ranganath, C. (2007). Reinforcement learning signals predict future decisions. *Journal of Neuroscience*, *27*(2). https://doi.org/10.1523/JNEUROSCI.4421-06.2007

Cona, G., & Semenza, C. (2017). Supplementary motor area as key structure for domain-general sequence processing: A unified account. In *Neuroscience and Biobehavioral Reviews* (Vol. 72). https://doi.org/10.1016/j.neubiorev.2016.10.033

Cover, T. M., & Thomas, J. A. (2005). Elements of Information Theory. In *Elements of Information Theory*. https://doi.org/10.1002/047174882X

Crinion, J. T., Warburton, E. A., Lambon-Ralph, M. A., Howard, D., & Wise, R. J. S. (2006). Listening to narrative speech after aphasic stroke: The role of the left anterior temporal lobe. *Cerebral Cortex*, *16*(8). https://doi.org/10.1093/cercor/bhj053

Crone, E. A., & Dahl, R. E. (2012). Understanding adolescence as a period of social-affective engagement and goal flexibility. *Nature Reviews Neuroscience*, *13*(9). https://doi.org/10.1038/nrn3313

De Lange, F. P., Hagoort, P., & Toni, I. (2005). Neural topography and content of movement representations. *Journal of Cognitive Neuroscience*, *17*(1). https://doi.org/10.1162/0898929052880039

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., Von Cramon, D. Y., & Engel, A. K. (2005). Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *Journal of Neuroscience*, *25*(50). https://doi.org/10.1523/JNEUROSCI.3286-05.2005

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. In *Trends in Cognitive Sciences* (Vol. 15, Issue 6). https://doi.org/10.1016/j.tics.2011.04.003

Denervaud, S., Gentaz, E., Matusz, P. J., & Murray, M. M. (2020). Multisensory Gains in Simple Detection Predict Global Cognition in Schoolchildren. *Scientific Reports*, *10*(1), 1394. https://doi.org/10.1038/s41598-020-58329-4

Desjardins, R. N., Rogers, J., & Werker, J. F. (1997). An exploration of why preschoolers perform differently than do adults in audiovisual speech perception tasks. *Journal of Experimental Child Psychology*, *66*(1), 85–110. https://doi.org/10.1006/jecp.1997.2379

Dionne-Dostie, E., Paquette, N., Lassonde, M., & Gallagher, A. (2015). Multisensory Integration and Child Neurodevelopment. *Brain Sciences*, *5*(1), 32–57. https://doi.org/10.3390/brainsci5010032

Diveica, V., Riedel, M. C., Salo, T., Laird, A. R., Jackson, R. L., & Binney, R. J. (2023). Graded functional organization in the left inferior frontal gyrus: evidence from task-free and task-based functional connectivity. *Cerebral Cortex*, *33*(23). https://doi.org/10.1093/cercor/bhad373

Döpfner, M., Plück, J., Kinnen, C., & Checklist, A. D. C. B. (2014). CBCL Handbuch-Schulalter. Manual zum Elternfragebogen über das Verhalten von Kindern und Jugendlichen. In *(CBCL/6-18R), zum Lehrerfragebogen über das Verhalten von Kindern und Jugendlichen (TRF/6-18R) und zum Fragebogen für Jugendliche (YSR/11-18R)*. Hogrefe Göttingen.

Eichenbaum, H. (2017). Memory: Organization and Control. *Annual Review of Psychology*, *68*. https://doi.org/10.1146/annurev-psych-010416-044131

Farrell, S., & Lewandowsky, S. (2018). Computational modeling of cognition and behavior. In *Computational Modeling of Cognition and Behavior*. https://doi.org/10.1017/CBO9781316272503

Fouragnan, E., Retzler, C., & Philiastides, M. G. (2018). Separate neural representations of prediction error valence and surprise: Evidence from an fMRI meta-analysis. *Human Brain Mapping*, *39*(7). https://doi.org/10.1002/hbm.24047

Fraga-González, G., Haller, P., Willinger, D., Gehrig, V., Frei, N., & Brem, S. (2025). Functional brain activations correlated with association strength and prediction error during novel symbol-speech sound learning. *Imaging Neuroscience*. https://doi.org/10.1162/imag_a_00439

Frei, N., Willinger, D., Haller, P., Fraga-González, G., Pamplona, G. S. P., Haugg, A., Lutz, C. G., Coraj, S., Hefti, E., & Brem, S. (2025). Towards a mechanistic understanding of reading difficulties: Deviant audiovisual learning dynamics and network connectivity in children with poor reading skills. *The Journal of Neuroscience*, e1119242025. https://doi.org/10.1523/JNEUROSCI.1119-24.2025

Friedrich, J., & Beste, C. (2019). The impact of stimulus modality on the processing of conflicting sensory information during response inhibition. *Neuroscience*, *410*. https://doi.org/10.1016/j.neuroscience.2019.05.010

Gao, C., Green, J. J., Yang, X., Oh, S., Kim, J., & Shinkareva, S. V. (2023). Audiovisual integration in the human brain: a coordinate-based meta-analysis. *Cerebral Cortex*, *33*(9). https://doi.org/10.1093/cercor/bhac443

Gao, C., & Shinkareva, S. V. (2021). Modality-general and modality-specific audiovisual valence processing. *Cortex*, *138*. https://doi.org/10.1016/j.cortex.2021.01.022

Garrison, J., Erdeniz, B., & Done, J. (2013). Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. In *Neuroscience and Biobehavioral Reviews* (Vol. 37, Issue 7). https://doi.org/10.1016/j.neubiorev.2013.03.023

Gentile, G., Björnsdotter, M., Petkova, V. I., Abdulkarim, Z., & Ehrsson, H. H. (2015). Patterns of neural activity in the human ventral premotor cortex reflect a whole-body multisensory percept. *NeuroImage*, *109*. https://doi.org/10.1016/j.neuroimage.2015.01.008

Gentile, G., Guterstam, A., Brozzoli, C., & Henrik Ehrsson, H. (2013). Disintegration of multisensory signals from the real hand reduces default limb self-attribution: An fMRI study. *Journal of Neuroscience*, *33*(33). https://doi.org/10.1523/JNEUROSCI.1363-13.2013

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? In *Trends in Cognitive Sciences* (Vol. 10, Issue 6, pp. 278–285). https://doi.org/10.1016/j.tics.2006.04.008

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). https://doi.org/10.1016/j.cub.2008.04.036

Grob, A., & Hagmann-von Arx, P. (2018). *IDS-2-Intelligence and Development Scales-2. Intelligenz-und Entwicklungsskalen für Kinder und Jugendliche-Testkoffer*. Hogrefe.

Hauser, T. U., Iannaccone, R., Walitza, S., Brandeis, D., & Brem, S. (2015). Cognitive flexibility in adolescence: Neural and behavioral mechanisms of reward prediction error processing in adaptive decision making during development. *NeuroImage*, *104*. https://doi.org/10.1016/j.neuroimage.2014.09.018

Hester, R., Barre, N., Murphy, K., Silk, T. J., & Mattingley, J. B. (2008). Human medial frontal cortex activity predicts learning from errors. *Cerebral Cortex*, *18*(8). https://doi.org/10.1093/cercor/bhm219

Hirst, R. J., Whelan, R., Boyle, R., Setti, A., Knight, S., O'Connor, J., Williamson, W., McMorrow, J., Fagan, A. J., Meaney, J. F., Kenny, R. A., De Looze, C., & Newell, F. N. (2021). Gray matter volume in the right angular gyrus is associated with differential patterns of multisensory integration with aging. *Neurobiology of Aging*, *100*. https://doi.org/10.1016/j.neurobiolaging.2020.12.004

Icenogle, G., & Cauffman, E. (2021). Adolescent decision making: A decade in review. In *Journal of Research on Adolescence* (Vol. 31, Issue 4). https://doi.org/10.1111/jora.12608

Joseph, R. (1990). *The Temporal Lobes*. https://doi.org/10.1007/978-1-4757-5969-3_7

Karipidis, I. I., Pleisch, G., Brandeis, D., Roth, A., Röthlisberger, M., Schneebeli, M., Walitza, S., & Brem, S. (2018). Simulating reading acquisition: The link between reading outcome and multimodal brain signatures of letter–speech sound learning in prereaders. *Scientific Reports*, *8*(1), 7121.

Karipidis, I. I., Pleisch, G., Di Pietro, S. V., Fraga-González, G., & Brem, S. (2021). Developmental Trajectories of Letter and Speech Sound Integration During Reading Acquisition. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.750491

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13, 1–26). https://doi.org/10.18637/JSS.V082.I13

Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal Oscillations and Multisensory Interaction in Primary Auditory Cortex. *Neuron*, *53*(2), 279–292. https://doi.org/10.1016/j.neuron.2006.12.011

Lauzon, S., Abraham, A. E., Curcin, K., Butler, B. E., & Stevenson, R. A. (2022). The relationship between multisensory associative learning and multisensory integration. *Neuropsychologia*, *174*. https://doi.org/10.1016/j.neuropsychologia.2022.108336

Lenhard, A., Lenhard, W., Segerer, R., & Suggate, S. (2015). *Peabody Picture Vocabulary Test-4. Ausgabe: Deutsche Fassung*. Frankfurt am Main: Pearson Assessment.

Lenth, R. (2024). *emmeans: Estimated marginal means, aka least-squares means* (R package version 1.10. 2). https://CRAN.R-project.org/package=emmeans

Li, W., Qin, W., Liu, H., Fan, L., Wang, J., Jiang, T., & Yu, C. (2013). Subregions of the human superior frontal gyrus and their connections. *NeuroImage*, *78*. https://doi.org/10.1016/j.neuroimage.2013.04.011

Limanowski, J., & Blankenburg, F. (2016). Integration of visual and proprioceptive limb position information in human posterior parietal, premotor, and extrastriate cortex. *Journal of Neuroscience*, *36*(9). https://doi.org/10.1523/JNEUROSCI.3987-15.2016

Limanowski, J., & Blankenburg, F. (2017). Posterior parietal cortex evaluates visuoproprioceptive congruence based on brief visual information. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-16848-7

Liu, X., Hairston, J., Schrier, M., & Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: A meta-analysis of functional neuroimaging

studies. In *Neuroscience and Biobehavioral Reviews* (Vol. 35, Issue 5, pp. 1219–1236). https://doi.org/10.1016/j.neubiorev.2010.12.012

Loureiro, R. J. (2020). Decision making in adolescents: A multifaceted construct. In *Journal of Human Growth and Development* (Vol. 30, Issue 2). https://doi.org/10.7322/JHGD.V30.10362

Ludersdorfer, P., Price, C. J., Kawabata Duncan, K. J., DeDuck, K., Neufeld, N. H., & Seghier, M. L. (2019). Dissociating the functions of superior and inferior parts of the left ventral occipito-temporal cortex during visual word and object processing. *NeuroImage*, *199*. https://doi.org/10.1016/j.neuroimage.2019.06.003

MathWorks Inc. (2020). *MATLAB, Version 9.9.0 (R2020b)*.

MathWorks Inc. (2022). *MATLAB version: 9.13.0 (R2022b)*. The MathWorks Inc Natick, Massachusetts.

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience*, *MAY*. https://doi.org/10.3389/fnhum.2011.00039

Matsumoto, M., Matsumoto, K., Abe, H., & Tanaka, K. (2007). Medial prefrontal cell activity signaling prediction errors of action values. *Nature Neuroscience*, *10*(5). https://doi.org/10.1038/nn1890

Mazaika, P., Whitfield-Gabrieli, S., Reiss, A., & Glover, G. (2007). Artifact repair for fMRI data from high motion clinical subjects. *Human Brain Mapping*, *47*(58), 70231–70238.

McClure, S. M., Berns, G. S., & Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron*, *38*(2). https://doi.org/10.1016/S0896-6273(03)00154-5

Misra, H., Ikbal, S., Bourlard, H., & Hermansky, H. (2004). Spectral entropy based feature for robust ASR. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *1*. https://doi.org/10.1109/icassp.2004.1325955

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

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of Cue Integration in Human Navigation. *Current Biology*, *18*(9). https://doi.org/10.1016/j.cub.2008.04.021

Navarro, D. J., & Fuss, I. G. (2009). Fast and accurate calculations for first-passage times in Wiener diffusion models. In *Journal of Mathematical Psychology* (Vol. 53, Issue 4). https://doi.org/10.1016/j.jmp.2009.02.003

Niv, Y., & Schoenbaum, G. (2008). Dialogues on prediction errors. In *Trends in Cognitive Sciences* (Vol. 12, Issue 7). https://doi.org/10.1016/j.tics.2008.03.006

Noesselt, T., Rieger, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., Heinze, H. J., & Driver, J. (2007). Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *Journal of Neuroscience*, *27*(42), 11431–11441. https://doi.org/10.1523/JNEUROSCI.2252-07.2007

Obin, N., & Liuni, M. (2012). On the generalization of Shannon entropy for speech recognition. *2012 IEEE Workshop on Spoken Language Technology, SLT 2012 - Proceedings*. https://doi.org/10.1109/SLT.2012.6424204

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron*, *38*(2). https://doi.org/10.1016/S0896-6273(03)00169-7

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

O'Reilly, J. X. (2013). Making predictions in a changing world-inference, uncertainty, and learning. *Frontiers in Neuroscience*, *7 JUN*. https://doi.org/10.3389/fnins.2013.00105

Pagnoni, G., Zink, C. F., Montague, P. R., & Berns, G. S. (2002). Activity in human ventral striatum locked to errors of reward prediction. *Nature Neuroscience*, *5*(2). https://doi.org/10.1038/nn802

Palejwala, A. H., Dadario, N. B., Young, I. M., O'Connor, K., Briggs, R. G., Conner, A. K., O'Donoghue, D. L., & Sughrue, M. E. (2021). Anatomy and White Matter Connections of the Lingual Gyrus and Cuneus. *World Neurosurgery*, *151*. https://doi.org/10.1016/j.wneu.2021.04.050

Pasqualotto, A., Cochrane, A., Bavelier, D., & Altarelli, I. (2024). A novel task and methods to evaluate inter-individual variation in audio-visual associative learning. *Cognition*, *242*. https://doi.org/10.1016/j.cognition.2023.105658

Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*, *22*(2), 237–247. https://doi.org/10.1016/S0163-6383(99)00003-X

Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, *6*(2), 191–196. https://doi.org/10.1111/1467-7687.00271

Pedersen, M. L., Frank, M. J., & Biele, G. (2017). The drift diffusion model as the choice rule in reinforcement learning. *Psychonomic Bulletin and Review*, *24*(4). https://doi.org/10.3758/s13423-016-1199-y

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1). https://doi.org/10.3758/s13428-018-01193-y

Penny, W., Friston, K., Ashburner, J., Kiebel, S., & Nichols, T. (2007). Statistical Parametric Mapping: The Analysis of Functional Brain Images. In *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. https://doi.org/10.1016/B978-0-12-372560-8.X5000-1

Pernet, C. R. (2014). Misconceptions in the use of the General Linear Model applied to functional MRI: A tutorial for junior neuro-imagers. In *Frontiers in Neuroscience* (Issue 8 JAN). https://doi.org/10.3389/fnins.2014.00001

Pleisch, G., Karipidis, I. I., Brem, A., Röthlisberger, M., Roth, A., Brandeis, D., Walitza, S., & Brem, S. (2019). Simultaneous EEG and fMRI reveals stronger sensitivity to orthographic strings in the left occipito-temporal cortex of typical versus poor beginning readers. *Developmental Cognitive Neuroscience*, *40*. https://doi.org/10.1016/j.dcn.2019.100717

R Core Team, R. (2024). *R: A language and environment for statistical computing*. R foundation for statistical computing, Vienna, Austria. https://www.R-project.org/

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. In *Neural Computation* (Vol. 20, Issue 4). https://doi.org/10.1162/neco.2008.12-06-420

Ratcliff, R., & Rouder, J. N. (1998). Modeling Response Times for Two-Choice Decisions. *Psychological Science*, *9*(5). https://doi.org/10.1111/1467-9280.00067

Rescorla, R. A., & Wagner, A. R. (1972). A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement. *Clasical Conditioning II: Current Research and Theory*.

Richlan, F. (2019). The functional neuroanatomy of letter-speech sound integration and its relation to brain abnormalities in developmental dyslexia. In *Frontiers in Human Neuroscience* (Vol. 13). https://doi.org/10.3389/fnhum.2019.00021

Rolls, E. T. (2021). The temporal cortex. In *Brain Computations: What and How*. Oxford University Press.

Ruiz-Rizzo, A. L., Neitzel, J., Müller, H. J., Sorg, C., & Finke, K. (2018). Distinctive correspondence between separable visual attention functions and intrinsic brain networks. *Frontiers in Human Neuroscience*, *12*. https://doi.org/10.3389/fnhum.2018.00089

Rushworth, M. F. S. (2008). Intention, choice, and the medial frontal cortex. In *Annals of the New York Academy of Sciences* (Vol. 1124). https://doi.org/10.1196/annals.1440.014

Seghier, M. L. (2013). The angular gyrus: Multiple functions and multiple subdivisions. In *Neuroscientist* (Vol. 19, Issue 1). https://doi.org/10.1177/1073858412440596

Seitz, A. R., Kim, R., Van Wassenhove, V., & Shams, L. (2007). Simultaneous and independent acquisition of multisensory and unisensory associations. *Perception*, *36*(10). https://doi.org/10.1068/p5843

Smith, A. R., Steinberg, L., & Chein, J. (2014). The role of the anterior insula in adolescent decision making. In *Developmental Neuroscience* (Vol. 36, Issues 3–4). https://doi.org/10.1159/000358918

Sours, C., Raghavan, P., Medina, A. E., Roys, S., Jiang, L., Zhuo, J., & Gullapalli, R. P. (2017). Structural and Functional Integrity of the Intraparietal Sulcus in Moderate and Severe Traumatic Brain Injury. *Journal of Neurotrauma*, *34*(7). https://doi.org/10.1089/neu.2016.4570

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. In *Nature Reviews Neuroscience* (Vol. 9, Issue 4, pp. 255–266). https://doi.org/10.1038/nrn2331

Stein, B. E., Stanford, T. R., & Rowland, B. A. (2020). Multisensory integration and the society for neuroscience: Then and now. In *Journal of Neuroscience* (Vol. 40, Issue 1). https://doi.org/10.1523/JNEUROSCI.0737-19.2019

Subramanian, A., Chitlangia, S., & Baths, V. (2022). Reinforcement learning and its connections with neuroscience and psychology. *Neural Networks*, *145*. https://doi.org/10.1016/j.neunet.2021.10.003

Team Audacity. (2024). *Audacity (r): free audio editor and recorder [computer program]* (3.5.1). http://audacity.sourceforge.net/

Thesen, T., McDonald, C. R., Carlson, C., Doyle, W., Cash, S., Sherfey, J., Felsovalyi, O., Girard, H., Barr, W., Devinsky, O., Kuzniecky, R., & Halgren, E. (2012). Sequential then interactive processing of letters and words in the left fusiform gyrus. *Nature Communications*, *3*. https://doi.org/10.1038/ncomms2220

Tibon, R., Fuhrmann, D., Levy, D. A., Simons, J. S., & Henson, R. N. (2019). Multimodal integration and vividness in the angular gyrus during episodic encoding and retrieval. *Journal of Neuroscience*, *39*(22). https://doi.org/10.1523/JNEUROSCI.2102-18.2018

Tyll, S., Budinger, E., & Noesselt, T. (2011). Thalamic influences on multisensory integration. In *Communicative and Integrative Biology* (Vol. 4, Issue 4). https://doi.org/10.4161/cib.15222

Van Atteveldt, N., Murray, M. M., Thut, G., & Schroeder, C. E. (2014). Multisensory integration: Flexible use of general operations. In *Neuron* (Vol. 81, Issue 6). https://doi.org/10.1016/j.neuron.2014.02.044

Van Leijenhorst, L., Zanolie, K., Van Meel, C. S., Westenberg, P. M., Rombouts, S. A. R. B., & Crone, E. A. (2010). What motivates the adolescent? brain regions mediating reward sensitivity across adolescence. *Cerebral Cortex*, *20*(1). https://doi.org/10.1093/cercor/bhp078

Vildavski, V., Lo Verde, L., Blumberg, G., Parsey, J., & Norcia, A. (2021). PseudoSloan: A perimetric-complexity and area-controlled font for vision and reading research. *Journal of Vision*, *21*(9). https://doi.org/10.1167/jov.21.9.2857

Wallace, M. T., Woynaroski, T. G., & Stevenson, R. A. (2020). Multisensory integration as a window into orderly and disrupted cognition and communication. In *Annual Review of Psychology* (Vol. 71). https://doi.org/10.1146/annurev-psych-010419-051112

Wang, F., Karipidis, I. I., Pleisch, G., Fraga-González, G., & Brem, S. (2020). Development of Print-Speech Integration in the Brain of Beginning Readers With Varying Reading Skills. *Frontiers in Human Neuroscience*, *14*(289). https://doi.org/10.3389/fnhum.2020.00289

Wechsler, D. (2012). Wechsler preschool and primary scale of intelligence—fourth edition. *The Psychological Corporation San Antonio, TX*.

Wechsler, D. (2017). Wechsler Intelligence Scale for Children–Fifth Edition (WISC-V)(dt. Bearbeitung durch F. Petermann). In *Frankfurt a. M.: Pearson Assessment*.

Wilke, M., Holland, S. K., Altaye, M., & Gaser, C. (2008). Template-O-Matic: A toolbox for creating customized pediatric templates. *NeuroImage*, *41*(3). https://doi.org/10.1016/j.neuroimage.2008.02.056

Wilson, R. C., & Collins, A. G. E. (2019). Ten simple rules for the computational modeling of behavioral data. *ELife*, *8*. https://doi.org/10.7554/eLife.49547

Wilson, R. P., Colizzi, M., Bossong, M. G., Allen, P., Kempton, M., MTAC, & Bhattacharyya, S. (2018). The Neural Substrate of Reward Anticipation in Health: A Meta-Analysis of fMRI Findings in the Monetary Incentive Delay Task. *Neuropsychology Review*, *28*(4), 496–506. https://doi.org/10.1007/s11065-018-9385-5

Wu, X., Wang, T., Liu, C., Wu, T., Jiang, J., Zhou, D., & Zhou, J. (2017). Functions of learning rate in adaptive reward learning. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00592

Yaple, Z. A., Yu, R., & Arsalidou, M. (2020). Spatial migration of human reward processing with functional development: Evidence from quantitative meta-analyses. *Human Brain Mapping*, *41*(14). https://doi.org/10.1002/hbm.25103

Zhang, L., Lengersdorff, L., Mikus, N., Gläscher, J., & Lamm, C. (2020). Using reinforcement learning models in social neuroscience: Frameworks, pitfalls and suggestions of best practices. *Social Cognitive and Affective Neuroscience*, *15*(6), 695–707. https://doi.org/10.1093/scan/nsaa089

Zimmermann, P., Fimm, B., & Gondan, M. (2002). *Testbatterie zur aufmerksamkeitsprüfung für kinder:(KiTAP);[das schloß der geister]*. Psytest.