

The role of general cognitive skills in integrating visual and linguistic information during sentence comprehension: Individual differences across the lifespan

Florian Hintz^{1,2}, Cesko C. Voeten^{3,4,5,6}, Dorottya Dobó^{7,8}, Krisztina Sára Lukics^{7,8} and Ágnes Lukács^{7,8}

¹*Philipps University of Marburg, Marburg, Germany*

²*Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*

³*University of Pennsylvania, Department of Linguistics, Philadelphia PA, United States*

⁴*University of Pennsylvania, Department of Biology, Philadelphia PA, United States*

⁵*University of Amsterdam, Amsterdam, The Netherlands*

⁶*Fryske Akademy, Leeuwarden, The Netherlands*

⁷*Department of Cognitive Science, Faculty of Natural Sciences, Budapest University of Technology and Economics, Budapest, Hungary*

⁸*MTA-BME Momentum Language Acquisition Research Group, Eötvös Loránd Research Network (ELKH), Budapest, Hungary*

Running title: Individual differences in language-vision interactions

Keywords: individual differences; language-vision interactions, processing speed, working memory

Correspondence should be addressed to:

Florian Hintz

Deutscher Sprachatlas

Philipps University of Marburg

Pilgrimstein 16

35032 Marburg (Germany)

florian.hintz@uni-marburg.de

Word count abstract (max. 200): 197

Word count body (max 4500 for Intro, Results, Discussion): 4210

Abstract

Individuals exhibit massive variability in general cognitive skills, which affect language processing. This variability is partly developmental. Here, we recruited a large sample of participants (N=487), ranging from 9 to 90 years of age, and examined the involvement of non-verbal processing speed (assessed using visual and auditory reaction time tasks) and working memory (assessed using forward and backward Digit Span tasks) in a visual world task. Participants saw two objects on the screen and heard a sentence that referred to one of them. In half of the sentences, the target object could be predicted based on verb-selectional restrictions. We observed evidence for anticipatory processing on predictable compared to non-predictable trials. Visual and auditory processing speed had main effects on sentence comprehension *and* facilitated predictive processing, as evidenced by an interaction. We observed only weak evidence for the involvement of working memory in predictive sentence comprehension. Age had a nonlinear main effect (younger adults responded faster than children and older adults), but it did not differentially modulate predictive and non-predictive processing, nor did it modulate the involvement of processing speed and working memory. Our results contribute to delineating the cognitive skills that are involved in language-vision interactions.

The contributions of general cognitive skills to integrating visual and linguistic information during sentence comprehension: Individual differences across the lifespan

Introduction

Spoken language comprehension occurs in vastly different contexts. Whereas unimodal contexts, such as listening to someone talk over the phone, entail the processing of a single input stream, multimodal contexts, such as being in a face-to-face conversation or comprehending speech with reference to the immediate visual surrounding, require the coordination of different input streams. That is, in such contexts, listeners must quickly integrate the information derived from processing the spoken input with information derived from visual processing (e.g., the interlocutor's facial movements¹ or properties of the objects in the visual world²). One notion that has gained considerable experimental support over the past 20 years is that listeners often predict words that are likely to come up next^{3,4}. Moreover, it has become clear that during language-vision interactions comprehenders exploit cues in both the spoken and the visual modality for generating predictions about upcoming words^{5,6}. The present study is concerned with individual variability in this behaviour. Specifically, we examined the contributions of general cognitive skills to integrating visual and linguistic information during sentence comprehension.

Studies investigating prediction during language-vision interactions have often used the visual world paradigm⁷. In this paradigm, participants are presented with displays featuring visual objects, typically two or four, and spoken language that is related to the visual input. In most studies, participants' eye movements are recorded and analysed. In a seminal study, Altmann and Kamide⁸ observed that participants looked at the picture of a cake already before it was mentioned, when listening to the sentence "The boy will eat the cake". These anticipatory eye movements most likely reflect that they predicted 'cake' as coming up next based on verb-selectional restrictions (the other co-present objects in the scene were inedible: train, ball, car).

Given the asserted role of predictive processing as a general organizing principle for human cognition^{9,10}, an important task for experimental research is to explore the limits of prediction and to capture variability between individuals in order to define the architecture and mechanisms underlying human information processing^{11,12}. With regards to predictive language processing, several visual

world studies examined the influence of linguistic and general cognitive person variables on language-mediated anticipatory eye movements. For example, in samples of university students, Hintz et al.¹³ and Rommers et al.¹⁴ observed that listeners with larger receptive vocabularies and better production fluency showed stronger evidence for prediction than individuals who scored lower on the respective tests. Similarly, Özkan et al.¹⁵ reported that measures of early language production abilities were positively associated with variance in eye-movement-based measures of prediction in children, aged between 4 and 8 years. In terms of general cognitive skills, Özkan and colleagues also found that children with better working memory abilities were more likely to engage in prediction than children with limited working memory. The latter finding aligns well with the results by Huettig and Janse¹⁶, who tested individuals aged between 32 and 77 years and observed facilitatory effects of working memory capacity and non-verbal processing speed: Larger working memory capacities and better non-verbal processing speed were associated with increased likelihood of anticipatory looks to upcoming targets.

Contributions of general cognitive skills, such as working memory (WM) and non-verbal processing speed (ProSpeed), to language comprehension are predicted by most contemporary accounts, as it has become consensus that language is not an isolated system in the human mind, but interfaces with other cognitive systems^{12,17}. However, such general cognitive skills (e.g., WM, ProSpeed) become particularly relevant during language-vision interactions, where language processing needs to be coordinated with visual processing.

Working memory, assumed to reflect the ability to hold a limited amount of information temporarily active^{18–20}, may – according to Huettig et al.²¹ – function during language-vision interactions as a nexus where linguistically-derived and visually-derived information are bound together and maintained active for shorter periods of time^{22,23}. WM capacity is often assessed using the Digit Span test^{24–26}, where participants are presented with sequences of spoken digits, which they are instructed to remember and subsequently reproduce in the order they were encountered (forward version) or in reverse order (backward version). Sequence length increases as participants respond correctly. While forward Digit Span gauges the storage of information in short-term memory, backward Digit Span additionally requires the manipulation of short-term memory contents.

ProSpeed has been linked to variability in many cognitive tasks^{27,28}. It is conceived of as an indicator for the speed with which individuals process different types of stimulus input and carry out mental operations. A popular way of assessing ProSpeed are reaction time tasks, where participants are presented with visual or auditory stimuli to which they need to respond as quickly as possible²⁹. During language-vision interactions, ProSpeed is likely to play an important role since these contexts require rapid linguistic (i.e., auditory) and visual processing, as well as the integration of both information streams.

There is plenty of evidence suggesting that individuals exhibit variability in WM and ProSpeed across the lifespan^{25,26,28,30–34}. The distribution in both skills across the lifespan is best captured by a U-shaped curve, where developing and aging individuals perform worse than younger and middle-aged adults. Although previous reports demonstrated that WM and ProSpeed affect predictive processing during language-vision interactions in both developing and ageing individuals^{15,16}, an important question that has not been addressed is whether the predictive power of these skills changes as a function of age. In other words, do WM and ProSpeed contribute to explaining variability in predictive processing during language-vision interactions differently in developing and ageing individuals? Moreover, while there is experimental evidence suggesting that WM and ProSpeed affect language-vision interactions, it is unclear at which level(s) precisely.

In the present study, we used an individual-differences approach to address these questions. We recruited a large number of healthy Hungarian participants, including children, younger and middle-aged adults, and ageing individuals (age ranged from 9 to 90 years). Given the vast variability in age (and associated cognitive abilities), tasks were needed that could be completed by all participants. Since Digit Span and reaction time tests have previously been used in these populations^{25,26,28,30–34}, we included forward and backward Digit Span and visual and auditory reaction time tests to assess WM and ProSpeed, respectively. In terms of measuring predictive processing in the visual world, we adapted a task previously developed by Hintz et al.³⁵, henceforth referred to as Predictive Processing in the Visual World (PPVW) task. Although Hintz and colleagues had administered this test in younger adults, aged between 18 and 30 years, it seemed likely that children and older adults would also be able to complete it (e.g., given its similarity to the task used in children as young 1 and 2 years

of age^{36,37}. Moreover, this task could be implemented as an online experiment straightforwardly and thus facilitated the recruitment of large numbers of participants required for an individual-differences approach. In this task, participants previewed a display of two objects, one on the left- and one on the right-hand side of the screen. Then, they were presented with a spoken sentence that referred to one of the objects. Sentences had the structure, i.e. Hungarian translation equivalent of, “The man yesterday *verbed* the *patient*”, where in half of the trials the patient could be predicted based on the verb-selectional restrictions (see Appendix for an item overview). Participants were instructed to press the button associated with the target as soon as they knew which of the two objects would be referred to. The reference point for the reaction times (RTs) was the target word onset.

Given the results by Hintz et al.³⁵, for predictable trials, we expected responses before spoken target onset or shortly thereafter as participants could predict which of the objects would be referred to. For non-predictable trials, we expected responses substantially later than target word onset, as the target objects could not be predicted before their onset and it takes approximately 430 ms to program and launch a button press for a speeded task that involves a choice component³⁵. The critical question was whether variability in the difference between predictable and non-predictable conditions was explained by the four individual-differences predictors (i.e., performance on visual/auditory RT tests, forward/backward Digit Span) and whether the difference varied as a function of age. The nature of the four individual differences predictors also enabled us to pinpoint at which levels these general cognitive skills may have their effect. That is, we may relate effects of visual and auditory ProSpeed to visual and spoken language processing, respectively. Moreover, our different measures of working memory enabled us to test whether, and if so to what extent, the ability to store information in short-term memory and apply additional operations to that contents affect the type of language-vision interactions tested here.

We used a statistical technique, quantile generalized additive mixed-effects models, that is suited for experiments where the dependent variable does not follow a normal distribution and where the effects of predictor variables may be nonlinear in nature (see Methods for more details). Moreover, in addition to examining the predictors’ effects on the median (a robust analogue to the mean for data with non-normal distributions) of the dependent variable as is done in most RT analyses

(but see Baayen & Smolka³⁸), we explored whether the predictors' effects differ when considering different portions of the data. Specifically, we additionally examined the predictors' effects on the 10% and the 90% quantiles of the RT distribution in the visual world task, which reflect fast responses (10% quantile) and slow responses (90% quantile), respectively.

Results

Table 1 lists the descriptive statistics for the PPVW task and the five predictor variables (age, visual/auditory reaction time, forward/backward DS. In the PPVW, the average RT in the non-predictable condition was 675 ms. In the predictable condition, participants responded substantially earlier, on average 114 ms after target word onset. Given that it takes approximately 430 ms to program and launch a button press for a speeded task that involves a choice component³⁵, these numbers suggest that participants prepared their response as soon as they recognized the unfolding target word in the non-predictable condition, but did so *before* spoken target onset in the predictable condition. The overall distribution of the RT data is visualised in Figure 1, which suggests a bimodal distribution with a substantial portion of RTs occurring before target onset (i.e., in the predictable condition) and another peak after target onset (mostly RTs belonging to the non-predictable condition). The three vertical lines in Figure 1 correspond to the three quantiles of the distribution targeted by our analyses. The green line corresponds to the 50% quantile (i.e., the overall median; analogous in interpretation to the mean known from traditional regression analysis). The red and blue lines correspond to the 10% and 90% quantiles of the distribution, respectively, reflecting fast and slow responses. The distribution plots in Figure 1 and the standard deviations in Table 1 suggest substantial variability in the RTs, especially in the predictable condition.

Table 1 and Figure 1

Mean RTs for the simple visual and auditory reaction time tasks were 351 ms and 334 ms. The average forward and backward DS were 6.52 and 5.30 digits. The descriptive statistics suggest massive variability between participants in all four tasks. The four panels in Figure 2 visualize the

relationships between age and the four individual-differences predictors by means of scatter plots. The fitted regression line in each plot suggests that, as expected, these relationships are best captured by U-shaped curves, illustrating that children and older adults performed poorer on the reaction time and working memory tests than younger adults did.

Figure 2

Table 2 lists the results of the parametric terms for the quantile GAMs. These are interpretable in the same way as coefficients from a linear regression. That is, the results indicate whether or not a parametric term significantly contributed to explaining variance in the dependent variable. Figure 3 plots the results for the smooth terms, i.e. Age and its interactions. These are marginal effects, i.e. they have the same interpretation as a regression coefficient, with the single difference being that the effect need not be (and indeed is not in our data) the same for all age values, but rather varies (smoothly) along the different age values. A visual heuristic for interpreting the plots in Figure 3 is to focus on the trajectories of the three lines (each referring to one of the quantile models). Each line is surrounded by its 95% credible interval; if this credible interval does not include zero, that part of the trajectory represents a statistically significant effect. Below, we describe the results individually for each of the three quantile models and discuss their commonalities and differences in the Discussion.

Table 2 & Figure 3

50% quantile model

The 50% quantile, i.e. the median RT, represents the midpoint of the data distribution and is quantile regression's closest analogue to the traditional linear regression. When the data deviate from normality, as ours do, the median provides a more robust estimate of the true overall effect.

In terms of results, we observed a strong effect of Condition: As was also suggested by the mean RTs, participants indeed responded significantly faster in the predictable condition than they did in the non-predictable condition. In addition, we observed main effects of visual and auditory RTs on

RTs in the PPVW task: The positive relationship suggested that participants with faster average responses on the visual and auditory RT tasks also responded faster on the PPVW task. Importantly, the effects of both individual-differences variables interacted with Condition such that faster average responses on the visual and auditory RT tasks were associated with faster RTs in the predictable compared to the non-predictable condition. We observed no significant main effects for either forward or backward DS, nor were these found to stand in significant interactions with Condition. While there was trend for an interaction between Condition and backward DS (better working memory skills associated with faster RTs in the predictable relative to non-predictable condition), there was no sign of an effect concerning forward DS.

The smooth terms (Figure 3) show that, as expected, the nonlinear distribution of participants' ages is represented well by a nonlinear spline. There was a significant main effect of Age on the overall RTs, highlighted in the left upper plot in Figure 3 by means of the red vertical lines: in between these (i.e., participants aged 19 to 25 years), participants were significantly faster by 56 to 66 ms compared to the average participant. None of the other smooths involving age reached significance anywhere along it.

10% quantile model

For the 10% quantile, corresponding to the fastest RTs in the distributions, we observed a main effect of Condition. The difference between predictable and non-predictable conditions was larger compared to the 50% quantile. The main effects of visual and auditory RT were similar to those in the 50% quantile model. Interestingly, while the effect of the visual RT interacted significantly with the effect of Condition (as it did in the 50% quantile: faster average RTs were associated with faster RTs in the predictable relative to the non-predictable condition), this was not the case for auditory RT. Similarly, we observed no evidence for main effects of forward or backward DS, nor interactions of these with Condition or an interaction between forward DS and Condition. There was a strong trend for an interaction between Condition and backward DS (better working memory skills associated with faster RTs in the predictable relative to the non-predictable condition).

The main smooth for age was again significant. Participants aged 20 to 24 years old responded significantly faster, by 51 to 55 ms, compared to the ‘average’ participant. As before, none of the other smooth terms, which included interactions of age with other predictors, showed a significant effect.

90% quantile model

As in the other two quantile models, we observed a main effect of Condition for the 90% quantile, which was, however, smaller compared to the other quantiles. As before, we observed main effects of visual and auditory RTs and no main effects of forward and backward DS. As in the other models, visual RT significantly interacted with Condition. In the 90% quantile model, we also observed evidence for a statistically robust interaction between Condition and backward DS (better working memory skills associated with faster RTs in the predictable relative to the non-predictable condition).

Concerning the smooth of age, children aged up to 11½ years were slower by 231 to 141 ms, and ages 19 to 26 were faster by 74 to 78 ms than the ‘average’ participant. As in the other quantiles, no other smooth terms than the main effect were significant.

Discussion

Using an individual-differences approach, the present study tested the contribution of general cognitive skills to the processing and coordination of visual and linguistic information during spoken sentence comprehension. We recruited a large sample of participants, which varied substantially in age, and asked them to complete tasks measuring visual and auditory processing speed as well as short-term and working memory. We observed that performance on these tasks co-varied with age, such that younger adults (i.e., from 18 up to their mid 20’s) performed better than children and older adults. We related participants’ performance on these tasks to their performance on a task measuring predictive and non-predictive sentence comprehension in the visual world and examined whether the explanatory power of the individual-differences variables for language-vision interactions changes across the lifespan. In doing so, we went beyond earlier studies^{15,16}, which either tested developing or

aging participants. The data were statistically analysed using quantile regression, which (a) accommodates nonlinear effects and (b) enables the exploration of different portions of the distribution of the data, which is not required to follow a normal distribution.

We observed robust evidence for condition differences across the three quantile models. That is, participants consistently responded earlier in the predictable than in the non-predictable condition when considering the 50%, the 10% and the 90% quantiles of the distribution. Although not the main focus, this finding corroborates earlier studies that used the visual world paradigm⁷ and demonstrates the converging validity of our adapted paradigm, which used reaction times instead eye movements as the dependent variable. In line with earlier research, our results thus show that developing and adult listeners readily integrate visual processing with spoken sentence comprehension to restrict the domain of subsequent reference^{8,16,36}.

Our analyses revealed that this behavior is to a considerable extent subserved by ProSpeed—a result well in line with those reported by Huettig and Janse¹⁶ and predicted by contemporary accounts of language processing^{12,17}. Unlike Huettig and Janse, who used principal component analysis to operationalize ProSpeed as a single variable based on multiple test scores, we submitted visual and auditory reaction time scores individually to our quantile regression analyses. In all three models, main effects of visual and auditory reaction time emerged, suggesting that participants who responded faster on the ProSpeed tasks also responded faster on the sentence comprehension task. Crucially, we also observed that visual and auditory ProSpeed interacted with Condition such that ProSpeed skills particularly benefitted processing in the predictable compared to the non-predictable condition. This finding is important since it highlights the specificity of the ProSpeed effects for the predictive processing. However, whereas visual reaction time showed this interaction across all three models, auditory reaction time showed this effect in the 50% but not in the 10% or 90% quantile models. Thus, whereas predictive language processing in the visual world involves both visual and auditory ProSpeed skills, *visual* ProSpeed in particular plays a robust role, being observed not only for RTs that are typical in magnitude, but across the whole board. One may speculate that this is the case because visual processing typically precedes auditory processing in the scenarios tested in the visual world paradigm. As in most earlier studies, we presented the two objects slightly before participants

heard the cues (i.e., the verbs) relevant for generating predictions about the upcoming targets. It is thus plausible that rapid visual processing (e.g., retrieving semantic information about the depicted objects such as verb-selectional restrictions) is particularly beneficial for an efficient mapping of vision-derived and language-derived information. Future research could follow up on this speculation by manipulating the time interval between display onset and onset of the relevant language cue to increase/decrease time pressure and/or by manipulating the visual set size (e.g., a larger number of objects may additionally increase the involvement of visual ProSpeed skills). Furthermore, future research could use a variety of visual and auditory ProSpeed tasks to further pinpoint the effects of ProSpeed on language-vision interactions.

Unlike in the study by Huettig and Janse¹⁶, we did not observe strong effects of WM. That is, our analyses did not reveal main effects of forward or backward DS, and the interaction between backward DS and Condition was only statistically reliable in the 90% quantile. One important difference between our study and that by Huettig and Janse is that our displays featured two instead of four objects. The demands on the WM system were therefore lower and although our participants differed massively in their WM capacities, these differences may not have affected performance on the visual world task due to the overall lower demands. It is interesting that larger WM capacity was associated with faster responding on the visual world task when considering the 90% quantile of the data distribution. Data points in the 90% quantile reflect the slowest of all responses. One interpretation is that WM positively affected these types of responses (e.g., lapses of attention) such that participants with better compared to worse WM abilities were faster to recover WM contents for their response.

The main focus of the present investigation was on the moderating effects of age. As expected, performance on the ProSpeed and WM tasks varied as a function of age in the form of a U-shaped curve: Young adults performed better than children, adolescents, middle-aged and older adults. Similarly, age also affected performance on the visual word task, with younger adults (~19-26 years of age) generally responding faster compared to the mean of all participants (i.e., the ‘average participant’). However, age did not interact significantly with Condition nor did any other model term involving age show an effect. In other words, while it was the case that younger adults were fastest to

carry out the visual world task, this was a general advantage and did not differentially affect predictive and non-predictive sentence comprehension. Moreover, the predictive power of visual and auditory RT tasks and forward back backward DS did *not* change significantly across the lifespan and their relative contribution to explaining variance in predictive and non-predictive sentence comprehension was comparable across age groups. One may have predicted a different pattern. For example, adults, especially older adults, have more language experience, resulting in more entrenched lexical representations^{39,40} and stronger associative connections between lexical representations compared to developing language users. Moreover, there is ample evidence for the decline of older adults' general cognitive skills (e.g., slower ProSpeed²⁸, reduced WM capacity^{26,30}). Against this background, one may have predicted that children rely more strongly on general cognitive skills than older adults to make up for the missing language experience. Older adults, on the other hand, may rely less strongly on general cognitive skills and more strongly on 'linguistic heuristics' (e.g., associative priming from predictive verbs to target objects may be more efficient in experienced than in developing language users). Instead, our results suggest that the extent to which ProSpeed and WM are involved in predictive sentence comprehension changes in step with age-related changes in these general cognitive skills.

At a more general level, our results provide support for previous experimental reports^{16,41} and contemporary accounts¹⁷ demonstrating the importance of ProSpeed for linguistic processing. As discussed above, we observed weaker effects of WM than Huettig and Janse¹⁶ did and weaker effects than predicted by the WM model by Huettig et al.²¹. Given the nature of language-vision interactions and its proposed underlying architecture and mechanisms, it is very likely that WM plays an important role in binding together visually-derived and language-derived information and in keeping these information bundles active throughout the trial. For the methodological reasons discussed above, it would therefore be premature to question WM's involvement in language-vision interactions. Note also that our data showed strong trends for the critical interactions between WM and Condition. An important methodological consideration for future research, in line with other considerations discussed above, is to increase the number of objects from two to three, or to increase

sentence complexity/difficulty to increase the demands on the WM system, which may elicit larger variability across participants.

To conclude, using an RT-based variant of the visual world paradigm and an individual-differences approach, we observed evidence for predictive sentence comprehension. Visual and auditory ProSpeed consistently contributed to explaining variance in this behaviour. The effects of WM in our data were restricted to a small portion of trials, subserving processing on ‘slow trials’. Participants’ ProSpeed and WM skills co-varied with their age. Age had an overall effect on carrying out the visual world sentence comprehension task, but it did not modulate the extent to which ProSpeed and WM are involved in predictive sentence comprehension. These results contribute to mapping out the cognitive skills that affect visual and spoken language processing in language-vision interactions, and inspire further research on the role of cognitive skills and language experience in age-related changes in language processing.

Methods

Participants

Child and adolescent participants were recruited from schools with the help of teachers. The majority of younger adult participants were recruited through university courses where they received credit for their participation. Adult participants were recruited through convenience sampling. Elderly participants (> 65 years) were recruited from retirement homes. Only participants with no history of neurodevelopmental disorders or cognitive impairment were included in the study. In the case of elderly participants, either the test results for the Mini Mental State Examination⁴² test were provided by the institutions or the Montreal Cognitive Assessment Test⁴³ was administered by an experimenter. Participants under 18 were tested with the informed consent of their parents. Participants between 14 and 18 also provided their own informed consent, as did all adult participants, in accordance with the principles set out in the Declaration of Helsinki. The study was approved by the United Ethical Review Committee for Research in Psychology (EPKEB-2018/87).

We collected data from 605 participants. During the pre-processing stages, 124 participants were excluded (see section Pre-Processing). Thus, data from 487 participants were included in the analysis. The mean age of these participants was 29 years (SD = 18, range = 9 - 90); 349 were female, 117 were male and 21 preferred not to say. Fifty-one participants were left-handed, 417 right-handed and 18 indicated not to have a dominant hand. At the time of testing, participants had received on average 13.74 years of schooling (including primary and secondary education, bachelor's, master's and PhD university programs; SD = 5.00, range = 2 - 42).

General procedure

The tests were administered as part of a larger test battery examining the relationships between language skills and other cognitive functions. All tests were programmed in PsychoPy⁴⁴; data were collected via the Pavlovia online platform (pavlovia.org). Auditory stimuli were presented over headphones. In the case of children and adolescents, tests were administered in small groups in a quiet classroom under the supervision of an experimenter. Younger and middle-aged adult participants completed the tests online from their homes. Elderly participants (> 65 years) were tested under the

supervision of an experimenter. The order of trials in each test was the same for each participant. In terms of test order, we provided participants with a suggestion, but they were free to choose a different order.

Test descriptions

Predictive processing in the visual world test (PPVW). We designed a Hungarian version of the Dutch equivalent of the predictive processing in the visual world test developed and piloted by Hintz et al.³⁵; it was similar to the task used by Mani and Huettig³⁶. Participants were presented with transitive spoken sentences describing simple events (Table 3, for examples of predictable and non-predictable items) and two pictures, arranged on left- and right-hand sides of the computer screen—one of which was the target. On non-predictable trials, both objects fitted the semantic restrictions of the verb (paint: snake, apple); on predictable trials, only one of the two objects fitted the restrictions of the verb (turn off: torch, apple). The test consisted of 47 items (23 non-predictable, 24 predictable, see Appendix, for all items). All sentences were spoken by a Hungarian female speaker and recorded in a quiet environment. The sound intensity of all sentence recordings was equalized. We used the same drawings as used by Hintz et al.³⁵, which were coloured versions of the drawings provided by Snodgrass and Vanderwaart⁴⁵.

Table 3

A trial proceeded as follows: A black fixation cross appeared in the middle of the screen against a white background for 1000 ms. Target and distractor objects were presented on the left and on the right of the screen. The position (left versus right) of the target versus the distractor object was counterbalanced across trials. The appearance of the two objects was synchronized to the onset of the spoken sentence; the target word was the final word in the sentence. We inserted the adverb ‘yesterday’ between agent and verb to provide participants with additional time to inspect the display and retrieve the object names. There were approximately 2.18 s (SD = 158 ms, range = 1.83–2.64 s) between sentence onset and target word onset. Participants were instructed to press the left (left-hand

object) or right (right-hand object) arrow on the keyboard to select the object that would be referred to in the sentence as soon as they knew. Their response terminated the trial and the next trial started.

Participants completed two practice trials (one non-predictable and one predictable) before the test phase (45 sentences, 22 non-predictable and 23 predictable). Presentation order of predictable and non-predictable trials was randomized. Response time (RT) was the difference between the onset of the target name and participants' button press.

Visual and auditory reaction time tests. To test visual and auditory processing speed, we used reaction time tasks that measured speed of processing in the visual and auditory modalities for simple stimuli. Visual and auditory reaction time tests were implemented as two tasks of the same test, presented in fixed order, starting with the visual task. In the visual task, an image of a red ball with white spots appeared in the middle of the screen as the target stimulus. In the acoustic task, a 440 Hz pure tone was presented. In the visual task, participants were instructed to look at a blank grey screen and monitor the appearance of the red ball in the middle of the screen, and to press the spacebar as fast as possible upon appearance. In the auditory task, participants were instructed to press the spacebar as soon as they heard the tone. Both tasks consisted of 32 trials. Inter-stimulus intervals varied randomly between 1223 and 4988 ms for the visual task, and between 1041 and 4883 ms for the auditory task. Participants' scores were the mean RTs in both task modalities.

Digit span tests. We used the forward (forward DS) and backward (backward DS) Digit Span tasks^{25,26} to measure short-term memory and working memory capacity. Forward and backward DS were administered as two tasks of the same test, starting with forward DS. Recordings of digits (from 1 to 9), spoken by a Hungarian male speaker, were presented in sequence. The of these recordings varied between 548 and 862 ms. Sound intensities of the recordings were equalized. In both tasks, participants were instructed to reproduce (by typing out the numbers on the keyboard) the presented digit sequences in the encountered order (forward DS task) or in the reverse order (backward DS task). Digits were presented with a stimulus onset asynchrony of 1000 ms. The sequence length varied between 2-10 digits, and sequences were presented in incremental order of length. Within each length,

participants were presented with four sequences; testing terminated after three incorrect responses within the same length. Forward and backward DS were operationalized as the longest sequence length with at least two correct responses.

Data analysis

Pre-processing. The dataset initially contained data from 605 participants. However, due to technical or human failures, some participants had no data for one of the four individual differences tasks. This concerned 97 participants, who were removed. Next, we excluded all participants ($n = 5$) with mean RTs larger than 1500 ms in either the visual or the auditory reaction time test. Since these simple responses can be carried out within 250 to 500 ms in healthy participants³⁵, average RTs beyond 1500 ms reflect that participants did not carry out the task seriously.

For the PPVW, we removed all participants who had negative mean RTs in the non-predictable condition. Since the verb semantics in the non-predictable condition did not provide information on which object would be referred to, responses before target onset were not possible. Eight participants consistently responded before target onset in the non-predictable condition, reflecting guessing or non-serious behaviour. These participants were removed. Furthermore, we removed all incorrect responses, remaining negative RTs in the non-predictable condition, and RTs shorter than -2000 ms and larger than +3000 ms. Following these exclusion processes, we checked how many trials each participant retained and excluded all participants with fewer than 34 trials in total (i.e., 75% of all trials), which led to further exclusion of eight participants. The statistical analysis was based on data from 487 participants who contributed 20,960 data points in total. Given 47 possible data points per participant (amounting to 22,889 in total), 1929 data points (8.43%) were removed during pre-processing.

Statistical modelling. Given the bimodal distribution of the RTs (Figure 1), we analysed the data using quantile regression. Quantile regression is a distribution-free method of regression analysis. That is, instead of assuming a Gaussian dependent variable and fitting its mean value, quantile regression fits a specific quantile of a dependent variable, which is not assumed to follow a particular

distribution. Fitting models for multiple quantiles enables us to examine the effects of our predictors (i.e., condition, age, visual/auditory RT, forward/backward DS) across the whole distribution of the data. That is, we can independently consider the median RT (the quantile regression analogue to a linear regression analysis), fast RTs (which represent the first mode in our data), and positive RTs (corresponding to the second mode). Our analytical approach is couched within the framework of *quantile generalized additive models*^{46,47}, which enables the inclusion of random effects (e.g., participants) and penalized splines (in these data, age, which requires a more flexible predictor than a single straight line, and trial number) in the model.

We ran three quantile models. The first model targeted the 50% quantile of the data distribution, which corresponds to the median, which – if the data were normally distributed – corresponds to the overall mean as typically modelled in standard (regression) analyses. Furthermore, we ran a model targeting the 10% quantile, which corresponded closely to the first peak in our bimodal data distribution, and we ran a model targeting the 90% quantile of the data distribution, which corresponded closely to the second peak in our data (see Figure 1, for an overview of the data distribution and the three quantiles). In all three cases, the models were fitted to all data, in each case determining regression coefficients that produced the smallest residuals given the requested quantile.

Each model had the same dependent variable – RT – and the same predictors. These were fixed effects for Condition (predictable vs. non-predictable, sum-coded), forward DS (mean-centred), backward DS (mean-centred), visual RT (mean-centred), and auditory RT (mean-centred). In addition, interactions were included between Condition and the latter four individual-differences measures. Furthermore, we added a smooth term for age, defined as a thin-plate regression spline with maximally 15 basis functions. Within the generalized-additive-modelling framework, smoothing splines are used to fit predictors nonlinearly, with the amount of deviation from a straight line determined automatically from the data in a way that balances between overfitting and underfitting (here, based on the REML criterion). This allows the models to account for non-uniform effects of age in our dataset on the dependent variable. We additionally included terms for Condition, the four individual-differences measures, and the interactions between these, which we also allowed to vary nonlinearly with age using an approach called ‘varying-coefficient regression’⁴⁸. The resulting nine

difference smooths, relative to the main effect of Age, were given a point constraint passing through the origin, such that their intercepts would not compete with the parametric terms for the same nine predictors. In terms of random effects, we added random intercepts by participants, random slopes for Condition by participants¹, and a smooth term for the trial number with appropriate penalties to the null space and the first derivative for this to be analogous in interpretation to a random factor. The latter was included to capture potential effects of fatigue, which may stack up in a nonlinear way as the experiment progresses.

The models as described above were fitted to the three quantiles using function `mqqgam` from R package `qgam`⁴⁹. Estimation relied on the Extended Fellner-Schall optimizer, implemented by Wood and Fasiolo⁵⁰. For the fixed effects in the models, significance of the results was assessed using Wald z-scores. For the smooth terms along age, we predicted each marginal smooth's linear-predictor matrix onto a 100-point mesh, spanning the lowest to highest age values in our data. Multiplying these linear predictors by the model coefficients gave the fitted trajectory of the smooth. We assessed which parts of this trajectory were significant on the basis of 95% Bayesian credible intervals, computed using the approach by Wood⁴⁸ (p. 293-294). We did not assess the random effects nor the smooth along trial numbers, since these nuisance terms were not of interest. We used function `gam.check` to confirm that the three models converged successfully and that all smooth terms had sufficient numbers of basis functions.

¹ While random effects are normally incompatible with quantile GAMs⁴⁷, we were able to fit them as penalized parametric terms using the `paraPen` argument to `gam`, which is equivalent.

Acknowledgements

This work was supported by the Momentum Research Grant of the Hungarian Academy of Sciences (Momentum 96233: “Profiling Learning Mechanisms and Learners: Individual Differences From Impairments to Excellence in Statistical Learning and in Language Acquisition” awarded to AL) and by the Netherlands Organization for Scientific Research (NWO), Gravitation grant ‘Language in Interaction’ (grant number 024.001.006).

Data availability statement (mandatory)

The raw data, analysis script, and item list are publicly available on OSF (<https://osf.io/wy3bv/>).

Author contributions

F.H.: conceptualization, writing; *C.C.V.*: data analysis, writing; *D. D.*: methodology; *K.S.L.*: methodology; writing; *A.L.*: conceptualization; methodology; writing

Competing interests

The authors declare not to have competing interests.

References

1. Holler, J. & Levinson, S. C. Multimodal language processing in human communication. *Trends Cogn. Sci.* **23**, 639–652 (2019).
2. Huettig, F., Mishra, R. K. & Olivers, C. N. Mechanisms and representations of language-mediated visual attention. *Front Psychol* **2**, 394 (2012).
3. Federmeier, K. D. Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology* **44**, 491–505 (2007).
4. Pickering, M. J. & Gambi, C. Predicting while comprehending language: A theory and review. *Psychol. Bull.* **144**, 1002–1044 (2018).
5. Ferreira, F., Foucart, A. & Engelhardt, P. E. Language processing in the visual world: Effects of preview, visual complexity, and prediction. *J. Mem. Lang.* **69**, 165–182 (2013).
6. Hintz, F., Meyer, A. S. & Huettig, F. Visual context constrains language-mediated anticipatory eye movements. *Q. J. Exp. Psychol.* **73**, 458–467 (2020).
7. Huettig, F., Rommers, J. & Meyer, A. S. Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychol. (Amst.)* **137**, 151–171 (2011).
8. Altmann, G. T. M. & Kamide, Y. Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* **73**, 247–264 (1999).
9. Bar, M. The proactive brain: using analogies and associations to generate predictions. *Trends Cogn. Sci.* **11**, 280–289 (2007).
10. Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* **36**, 181–204 (2013).
11. Huettig, F. & Mani, N. Is prediction necessary to understand language? Probably not. *Lang. Cogn. Neurosci.* **31**, 19–31 (2016).
12. Kidd, E., Donnelly, S. & Christiansen, M. H. Individual Differences in Language Acquisition and Processing. *Trends Cogn. Sci.* **22**, 154–169 (2018).
13. Hintz, F., Meyer, A. S. & Huettig, F. Predictors of verb-mediated anticipatory eye movements in the visual world. *J. Exp. Psychol. Learn. Mem. Cogn.* **43**, 1352–1374 (2017).

14. Rommers, J., Meyer, A. S. & Huettig, F. Verbal and nonverbal predictors of language-mediated anticipatory eye movements. *Atten. Percept. Psychophys.* **77**, 720–730 (2015).
15. Özkan, D., Küntay, A. C. & Brouwer, S. The role of verbal and working memory skills in Turkish-speaking children's morphosyntactic prediction. *Appl. Psycholinguist.* **43**, 1305–1328 (2022).
16. Huettig, F. & Janse, E. Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Lang. Cogn. Neurosci.* **31**, 80–93 (2016).
17. McQueen, J. M. & Meyer, A. S. Key issues and future directions: Towards a comprehensive cognitive architecture for language use. in *Human language: From genes and brain to behavior* (ed. Hagoort, P.) 85–96 (MIT Press., 2019).
18. Baddeley, A. D. & Hitch, G. Working Memory. *Psychol. Learn. Motiv.* **8**, 47–89 (1974).
19. Baddeley, A. Working Memory. *Science* **255**, 556–559 (1992).
20. Baddeley, A. The episodic buffer: a new component of working memory? *Trends Cogn. Sci.* **4**, 417–423 (2000).
21. Huettig, F., Olivers, C. N. L. & Hartsuiker, R. J. Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychol. (Amst.)* **137**, 138–150 (2011).
22. Knoeferle, P. & Crocker, M. W. The influence of recent scene events on spoken comprehension: Evidence from eye movements. *J. Mem. Lang.* **57**, 519–543 (2007).
23. Spivey, M. J., Richardson, D. C. & Fitneva, S. A. Thinking outside the brain: Spatial indices to visual and linguistic information. in *The interface of language, vision, and action: Eye movements and the visual world* (eds. Henderson, J. & Ferreira, F.) 161–190 (Academic Press, 2004).
24. Alan D. Baddeley. *Working memory*. (1986).
25. Wechsler, D. *Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV)*. (APA PsycTests, 2003).
26. Wechsler, D. *Wechsler Adult Intelligence Scale, Fourth Edition (WAIS-IV)*. (APA PsycTests, 2008).

27. Schubert, A.-L., Hagemann, D., Voss, A., Schankin, A. & Bergmann, K. Decomposing the relationship between mental speed and mental abilities. *Intelligence* **51**, 28–46 (2015).
28. Salthouse, T. A. The processing-speed theory of adult age differences in cognition. *Psychol. Rev.* **103**, 403–428 (1996).
29. Deary, I. J., Liewald, D. & Nissan, J. A free, easy-to-use, computer-based simple and four-choice reaction time programme: The Deary-Liewald reaction time task. *Behav. Res. Methods* **43**, 258–268 (2011).
30. Bopp, K. L. & Verhaeghen, P. Aging and Verbal Memory Span: A Meta-Analysis. *J. Gerontol. Ser. B* **60**, P223–P233 (2005).
31. Fry, A. F. & Hale, S. Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychol. Sci.* **7**, 237–241 (1996).
32. Gathercole, S. & Alloway, T. P. *Working memory and learning: A practical guide for teachers.* (Sage, 2008).
33. Kail, R. Developmental change in speed of processing during childhood and adolescence. *Psychol. Bull.* **109**, 490–501 (1991).
34. Salthouse, T. A. *Theoretical perspectives on cognitive aging.* (Psychology Press, 1991).
35. Hintz, F., Dijkhuis, M., van ‘t Hoff, V., McQueen, J. M. & Meyer, A. S. A behavioural dataset for studying individual differences in language skills. *Sci. Data* **7**, 429 (2020).
36. Mani, N. & Huettig, F. Prediction during language processing is a piece of cake—But only for skilled producers. *J. Exp. Psychol. Hum. Percept. Perform.* **38**, 843–847 (2012).
37. Reuter, T., Mazzei, C., Lew-Williams, C. & Emberson, L. Infants’ lexical comprehension and lexical anticipation abilities are closely linked in early language development. *Infancy* (2023) doi:<https://doi.org/10.1111/infa.12534>.
38. Baayen, R. H. & Smolka, E. Modeling morphological priming in German with naive discriminative learning. *Front. Commun.* **5**, 17 (2020).
39. Diependaele, K., Lemhöfer, K. & Brysbaert, M. The word frequency effect in first-and second-language word recognition: A lexical entrenchment account. *Q. J. Exp. Psychol.* **66**, 843–863 (2013).

40. Keuleers, E., Stevens, M., Mandera, P. & Brysbaert, M. Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *Q. J. Exp. Psychol.* **68**, 1665–1692 (2015).
41. Hintz, F. *et al.* Shared lexical access processes in speaking and listening? An individual differences study. *J. Exp. Psychol. Learn. Mem. Cogn.* **46**, 1048–1063 (2020).
42. Folstein, M. F., Robins, L. N. & Helzer, J. E. The mini-mental state examination. *Arch. Gen. Psychiatry* **40**, 812–812 (1983).
43. Volosin, M., Janacsek, K. & Németh, D. Hungarian version of the Montreal Cognitive Assessment (MoCA) for screening mild cognitive impairment. *Psychiatr. Hung. Magy. Pszichiatriai Tarsasag Tudomanyos Folyoirata* **28**, 370–392 (2013).
44. Peirce, J. *et al.* PsychoPy2: Experiments in behavior made easy. *Behav. Res. Methods* **51**, 195–203 (2019).
45. Snodgrass, J. G. & Vanderwart, M. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *J. Exp. Psychol. [Hum. Learn.]* **6**, 174–215 (1980).
46. Fasiolo, M., Wood, S. N., Zaffran, M., Nedellec, R. & Goude, Y. qgam: Bayesian Nonparametric Quantile Regression Modeling in R. *J. Stat. Softw.* **100**, 1–31 (2021).
47. Hintz, F., Voeten, C. C. & Scharenborg, O. Recognizing non-native spoken words in background noise increases interference from the native language. *Psychon. Bull. Rev.* (2022) doi:10.3758/s13423-022-02233-7.
48. Wood, S. N. *Generalized additive models: an introduction with R.* (CRC press, 2017).
49. Fasiolo, M., Wood, S. N., Zaffran, M., Goude, Y. & Nedellec, R. qgam: Smooth Additive Quantile Regression Models. (2021).
50. Wood, S. N. & Fasiolo, M. A generalized Fellner-Schall method for smoothing parameter optimization with application to Tweedie location, scale and shape models. *Biometrics* **73**, 1071–1081 (2017).

Table 1. Descriptive statistics for dependent and independent variables for participants included in the analysis (n = 487).

Variable	M	SD	Range
PPVW (ms)			
<i>Predictable</i>	114	553	-1290 – 2990
<i>Non-predictable</i>	675	300	202 – 2929
Age (years)	29.25	17.78	8.70 – 89.60
Visual RT (ms)	351	132	231 – 1219
Auditory RT (ms)	334	83	197 – 801
Forward DS (score)	6.52	1.76	2 – 10
Backward DS (score)	5.30	1.97	2 – 10

Table 2. Results from the three quantile GAMs. Estimates are on the ms scale.

Quantile = 50% (distribution median)			
Predictor	Estimate (SE)	z	p
Intercept	369.64 (12.13)	30.46	<.001***
Condition	-564.48 (40.22)	-14.04	<.001***
Forward DS	3.56 (7.19)	0.49	.62
Backward DS	-10.74 (8.46)	-1.27	.20
Visual RT	0.50 (0.13)	3.76	<.001***
Auditory RT	0.86 (0.26)	3.27	<.01**
Condition × Forward DS	5.07 (9.05)	0.56	.56
Condition × Backward DS	-15.25 (8.34)	-1.83	.07
Condition × Visual RT	0.40 (0.16)	2.57	.01*
Condition × Auditory RT	0.49 (0.23)	2.10	.04*
Quantile = 10% (fast responses)			
Predictor	Estimate (SE)	z	p
Intercept	58.59 (11.07)	5.30	<.001***
Condition	-724.72 (39.26)	-18.46	<.001***
Forward DS	6.95 (6.54)	1.06	.29
Backward DS	-10.24 (7.70)	-1.33	.18
Visual RT	0.50 (0.12)	4.17	<.001***
Auditory RT	0.60 (0.25)	2.40	.02*
Condition × Forward DS	7.97 (8.64)	0.92	.36
Condition × Backward DS	-15.51 (7.93)	-1.96	.05
Condition × Visual RT	0.54 (0.15)	3.64	<.001***
Condition × Auditory RT	0.17 (0.21)	0.84	.40
Quantile = 90% (slow responses)			
Predictor	Estimate (SE)	z	p
Intercept	757.53 (12.86)	58.92	<.001***
Condition	-358.17 (36.20)	-9.89	<.001***
Forward DS	0.36 (7.49)	0.05	.96
Backward DS	-11.21 (8.93)	-1.26	.21
Visual RT	0.47 (0.13)	3.46	<.001***
Auditory RT	0.96 (0.32)	2.95	<.01**
Condition × Forward DS	-2.68 (7.60)	-0.35	.72
Condition × Backward DS	-14.79 (6.77)	-2.19	.03*
Condition × Visual RT	0.62 (0.31)	2.00	.045*
Condition × Auditory RT	0.28 (0.17)	1.64	.10

Table 3: Stimulus examples for predictable and non-predictable items in the PPVW task.

Predictable

A férfi tegnapelőtt bekapcsolta az elemlámpát.

The man the day before yesterday turn PAST3SgDef the
torch-ACC

‘The man turned off the torch the day before yesterday’

Images: apple, torch

Non-predictable

A férfi tegnapelőtt lefestett egy kígyót.

The man the day before yesterday painted PAST3SgDef
a snake-ACC

‘The man painted a snake the day before yesterday’.

Images: snake, glasses

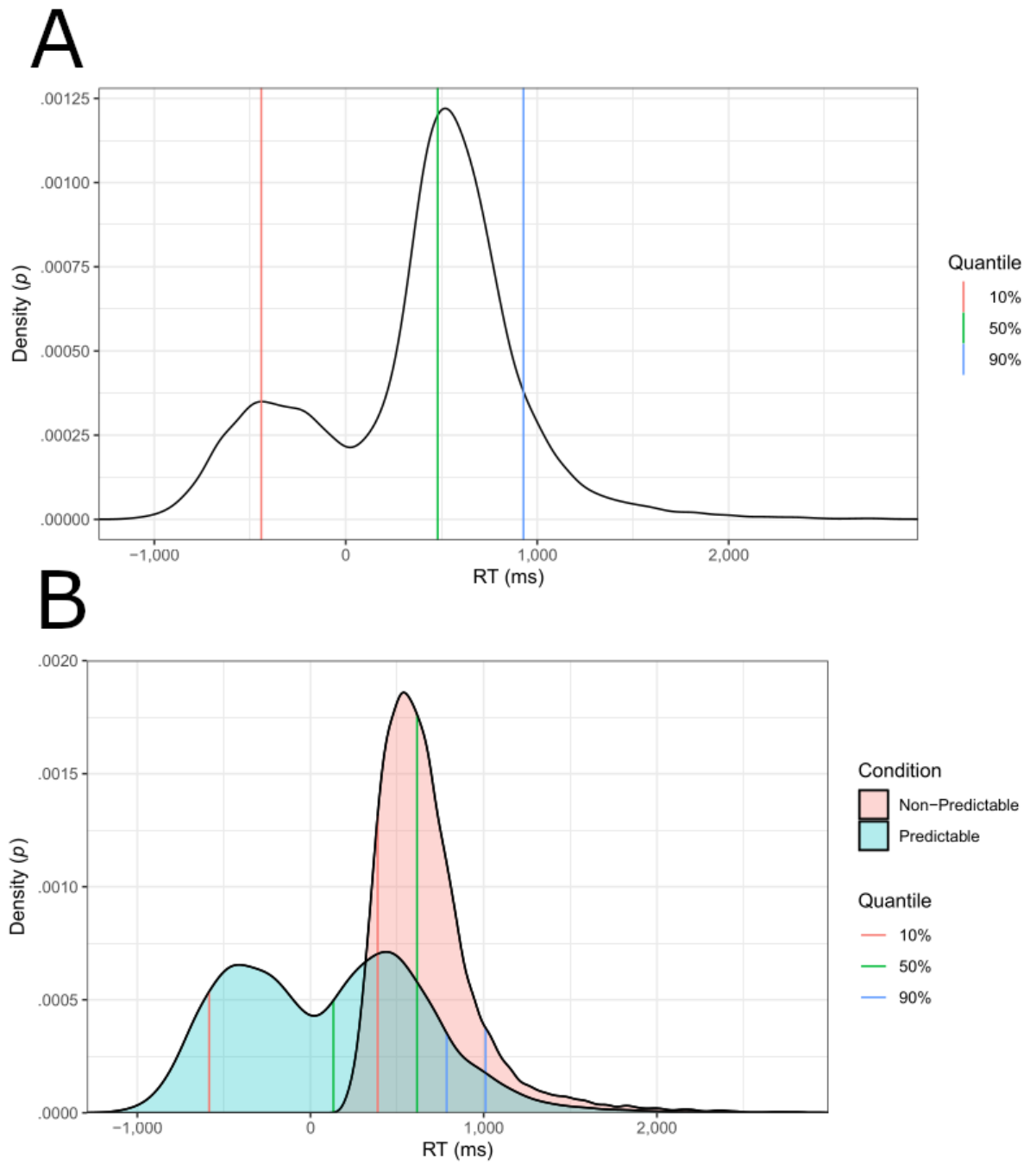


Figure 1: *Panel A.* Overall RT distribution in the PPVW task. Green (50%), red (10%) and blue (90%) vertical lines correspond to the quantiles targeted by our analyses. *Panel B.* RT distribution split out by condition (predictable vs. non-predictable).

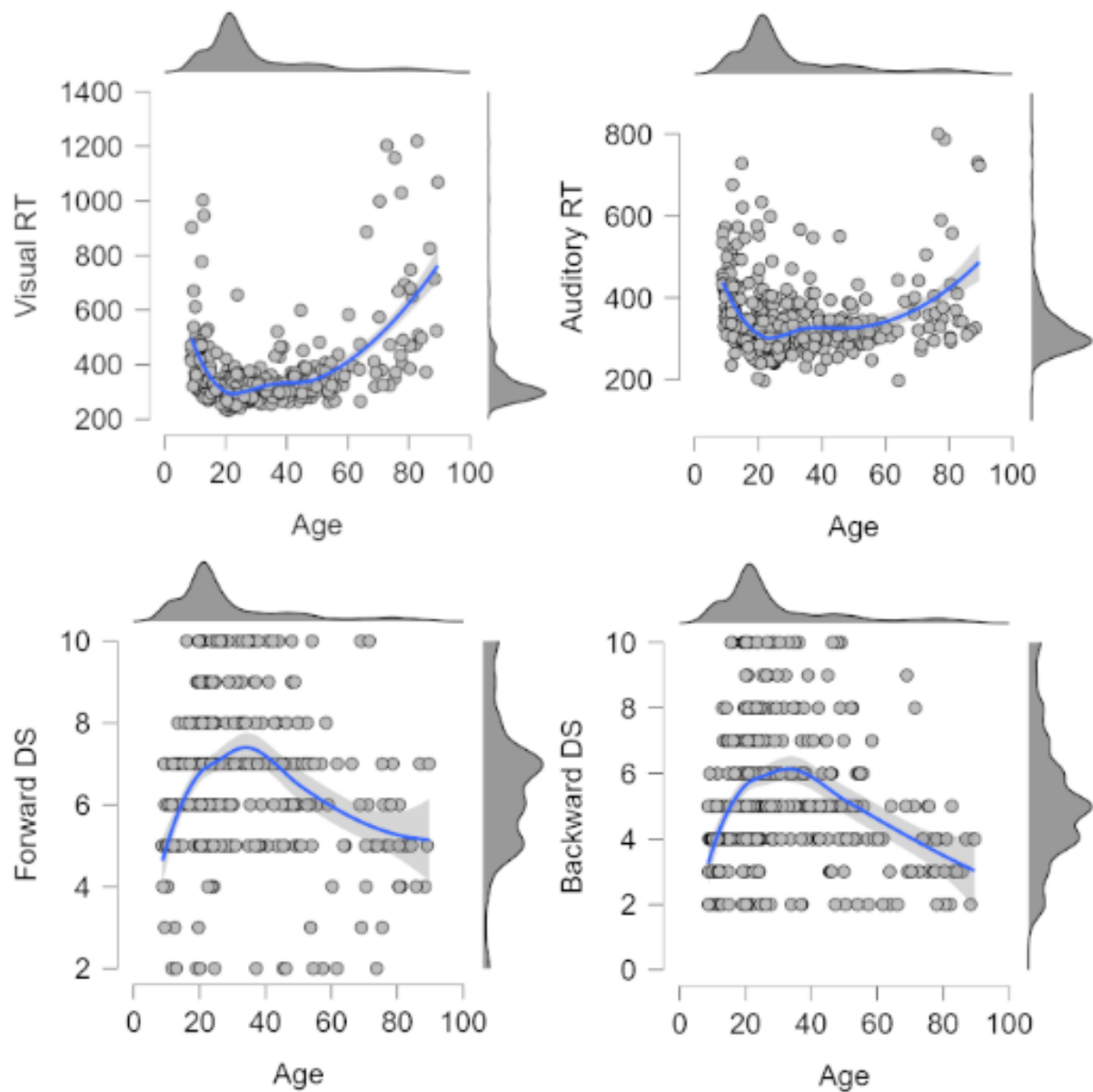


Figure 2: Scatter plots showing the relationships between age and the individual-differences predictors (visual/auditory reaction time, forward/backward Digit Span). Smoothed regression lines are accompanied by 95% credible intervals. Data distribution in each variable is visualized above and on the right of each scatter plot.

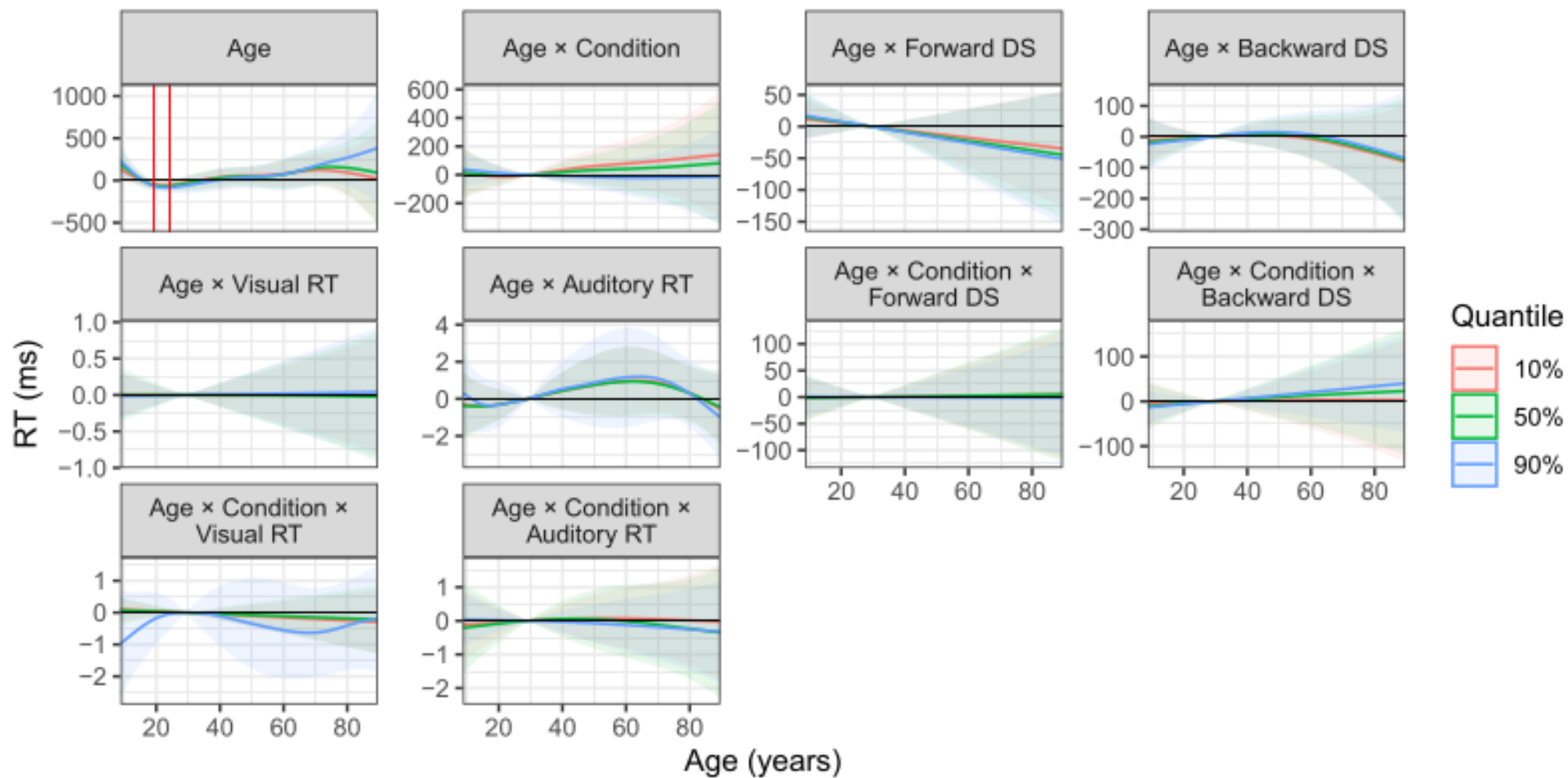


Figure 3. Results for the smooth terms in the quantile models.