Language, Cognition and Neuroscience

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/plcp21

Decomposing response times in Williams syndrome in separate cognitive processes

Dimitris Katsimpokis, Leendert van Maanen & Spyridoula Varlokosta

To cite this article: Dimitris Katsimpokis, Leendert van Maanen & Spyridoula Varlokosta (2022) Decomposing response times in Williams syndrome in separate cognitive processes, Language, Cognition and Neuroscience, 37:4, 440-456, DOI: 10.1080/23273798.2021.1985151

To link to this article: https://doi.org/10.1080/23273798.2021.1985151

LANGUAGE, COGNITION AND NEUROSCIENCE
2022, VOL. 37, NO. 4, 440–456
https://doi.org/10.1080/23273798.2021.1985151

Routledge
Taylor & Francis Group

REGULAR ARTICLE

Check for updates

Decomposing response times in Williams syndrome in separate cognitive processes

Dimitris Katsimpokis [a,b], Leendert van Maanen [a,c] and Spyridoula Varlokosta [d]

[a]Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands; [b]Department of Psychology, University of Basel, Basel, Switzerland; [c]Department of Experimental Psychology, Utrecht University, Utrecht, the Netherlands; [d]Department of Linguistics, Faculty of Philology, National and Kapodistrian University of Athens, Athens, Greece

ABSTRACT
Williams Syndrome (WS) is a rare neurodevelopmental disorder of genetic origin. The syndrome is characterised by a selective set of deficits in a number of cognitive domains. In spite of a wealth of studies, response times (RTs) of WS have attracted little attention. In the present study, we fill this gap by analysing data from a receptive vocabulary task using the Diffusion Decision Model (DDM). Our results show that the speed of accumulation, decision threshold and non-decision time parameters of WS individuals are similar to these of typically developing 5-year-old preschoolers. In addition, WS verbal intelligence scores were associated with the speed of accumulation of lexical information. Finally, the performance of WS and preschooler individuals was correlated across the vocabulary task and an additional orientation discrimination task only at the group but not at the individual level; therefore, pointing to domain-specific lexical and perceptual processing in WS.

ARTICLE HISTORY
Received 9 October 2020
Accepted 16 September 2021

KEYWORDS
Williams syndrome; diffusion decision model; receptive vocabulary; orientation discrimination; atypical development

1. Introduction

Over the last decades, a lively interest has sparked off in the cognitive profile of Williams Syndrome (WS). Interest in the syndrome coincided with a general trend in psychological and psycholinguistic research to consider atypical development as a fruitful empirical domain to test the assumptions and predictions of different cognitive theories (Karmiloff-Smith, 1998, 2009). In particular, WS has attracted substantial attention mainly because of its multifaceted cognitive profile. WS is a neurodevelopmental disorder caused by a deletion of 26–28 genes from chromosome 7, that involves cardiovascular, cognitive and neurophysiological deficits (Martens et al., 2008; Pober, 2010). Individuals with WS exhibit low general intelligence with mean IQ being equal to 55 (ranging from 40 to 90) corresponding to 4–6 year of typical development (Bellugi et al., 2000; Bennett et al., 1978; Martens et al., 2008); however, not all cognitive domains share the same degree of deficit. The present study addresses this complex array of cognitive strengths and weaknesses that characterise WS by using RTs and performance accuracy together in order to infer latent components of the Diffusion Decision Model (DDM) that relate to different aspects of WS lexical cognition.

Early linguistic studies noticed the unexpected fluency that characterises WS in terms of high competence in expressive and receptive vocabulary as well as grammatical comprehension (Bellugi et al., 1988). Later research, however, showed that WS individuals have a complex array of strengths and weaknesses characterising several cognitive domains of the syndrome (Bellugi et al., 2000; Brock, 2007; Landau & Ferrara, 2013). In the case of language, this was exemplified in processing asymmetries found across various levels of linguistic analysis (Brock, 2007; Camp & Karmiloff-Smith, 2015; Joffe & Varlokosta, 2007; Laing et al., 2001). For example, verbal memory span tasks reveal that, apart from chronologically delayed phonological similarity and word length effects, people with WS show word frequency effects much stronger than their typical developing counterparts, something which points towards a reduced contribution of the lexical/semantic level to the phonological processing (Vicari & Carlesimo, 2006). In addition, although WS grammatical comprehension of passive sentences or binding in simple (non-embedded) constructions is well developed (Clahsen & Almazan, 1998), other aspects of grammar such as morphological gender assignment appears deficient (Brock, 2007; Grant et al., 2002; Karmiloff-Smith et al., 1997). On

CONTACT Dimitris Katsimpokis dimitris.katsimpokis@unibas.ch Department of Psychology, University of Basel, Missionsstrasse 62a, Basel 4055, Switzerland
Supplemental data for this article can be accessed at https://doi.org/10.1080/23273798.2021.1985151

the other hand, WS receptive vocabulary is exceptionally well-developed as opposed to other aspects of vocabulary, and it sometimes surpasses the performance of typically developing children (Robinson et al., 2003). This line of evidence led researchers to suggest that individuals with WS mostly perform as well as it would be expected on the basis of their mental age but not as expected for their chronological age (Karmiloff-Smith et al., 1997, 2003; Landau & Ferrara, 2013).[1]

The multifaceted pattern of strengths and weaknesses appears in other cognitive domains in WS as well, such as motor and visuospatial processing. Motor development in WS is critically delayed (Carrasco et al., 2005; Hudson & Farran, 2011), with difficulties found in motor planning (Elliott et al., 2006) and execution (Krishnan et al., 2015; Withers, 1996). With respect to visuospatial processing, difficulties have been observed in orientation discrimination, object recognition, the approximate number system and geometric reorientation (Ansari et al., 2007; Lakusta et al., 2010; Landau et al., 2006; Libertus et al., 2014; Palomares et al., 2009). Some of these difficulties are reflected in the processing of spatial language (Laing & Jarrold, 2007; Landau & Hoffman, 2005). As in the case of language, however, some aspects of visuospatial processing are better developed than others, for example, integration of spatial information (Palomares et al., 2009) as well as visual recognition of objects when viewed in their canonical position (Landau et al., 2006).

1.1. Diffusion decision model

The DDM has originated from the family of sequential sampling models (Brown & Heathcote, 2008; Forstmann et al., 2016; Ratcliff, 1978; Ratcliff et al., 2016; Ratcliff & Smith, 2004; Stone, 1960). The DDM assumes that decisions are made based on the continuous stochastic accumulation of evidence. In the context of binary choice, decision-makers accumulate information for the two choices up to a point where one of them dominates. An example of an accumulation process is presented in Figure 1.

The DDM employs a certain set of parameters that are used to explain aspects of the accumulation process within a block of trials under the same experimental condition (Voss et al., 2004). The *drift-rate* parameter quantifies the strength of accumulation, namely how fast perceptual sampling informs either of two choices. In the case of receptive vocabulary, the drift-rate is taken to be an indication of internal memory activation for vocabulary items. The higher the drift-rate, the easier is retrieval from memory (Ratcliff, 1978; Ratcliff, Gomez, et al., 2004; van Maanen et al., 2012). Moreover, studies have found that higher drift-rates correlate with older ages (Ratcliff et al., 2012) and higher intelligence (Ratcliff et al., 2008; van Ravenzwaaij et al., 2011).

The *threshold* parameter encodes the amount of information required to commit to a decision. Thus, thresholds reflect the response strategy employed by participants. Higher thresholds indicate decision conservativeness or high response caution, whereas lower thresholds indicate less conservative decisions since not much evidence is required to commit to a decision. The threshold parameter has been found to indicate response caution in a number of psychophysical tasks (Bogacz et al., 2010; Forstmann et al., 2008; Miletić & van Maanen, 2019; van Maanen et al., 2011; van Maanen et al., 2016; Winkel et al., 2012), including lexical ones (Wagenmakers et al., 2008). Older individuals exhibit increased conservativeness as compared to children (Ratcliff et al., 2012) and conservativeness increases with age (Archambeau et al., 2020; Ratcliff et al., 2001, 2007; Ratcliff, Thapar, et al., 2004).

The DDM assumes that RTs come from a mixture of two time distributions: one related to accumulating evidence for the decision and another related to the time of sensory encoding and response execution. The parameter of *non-decision time,* which acts as an intercept[2], encodes when the earliest response is given by the participants, thus quantifying the time which is not related to evidence accumulation per se. Non-decision time is typically correlated with age (Ratcliff et al., 2012).[3]

Furthermore, a recent line of literature suggests that the DDM can inform controversial aspects of the behaviour of atypical populations other than WS. For example, Mulder et al. (2010) used the DDM in order to investigate how individuals with Attention Deficit Hyperactivity Disorder (ADHD) perform a perceptual decision-making task under different performance cues (i.e. emphasis on responding fast vs. accurately). Their results showed that ADHD participants scored lower per minute on average than a control group, when the speed performance cue is emphasised. By applying the DDM, they found that, although individuals with ADHD can trade-off speed over accuracy when asked to do so, they do not optimally execute this trade-off, reflected by suboptimal changes in the threshold parameter of the DDM, as compared to the control group. This was in spite of the fact that the ability of people with ADHD to collect perceptual information was comparable to that of the control group (as reflected in the DDM's drift-rate). These results reinforced the view that it is not cognition that is impaired in general in ADHD, but rather the aspect of behaviour optimisation (see also Pedersen et al., 2017). Other domains where the DDM has been applied to clinical populations include Obsessive

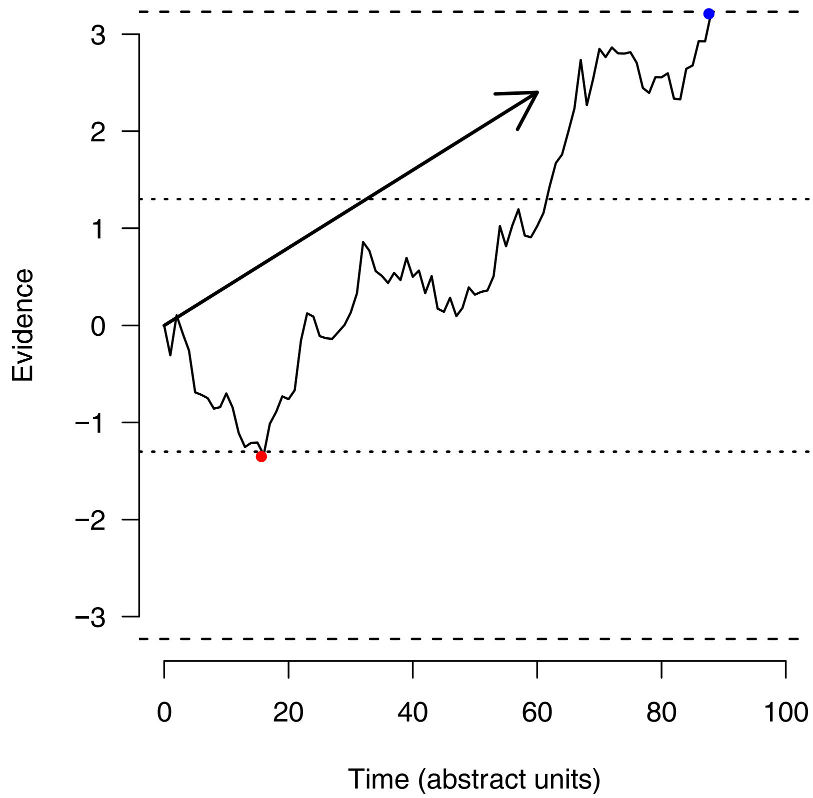

Figure 1. Example of an evidence accumulation process in a condition block. The stochastic line represents the accumulated evidence at every point in time. Correct decisions are made when the process hits the upper bound and error ones when hitting the lower bound. The dashed lines represent a conservative response caution strategy (i.e. thresholds farther away from the starting point) allowing more information to be accumulated so that the process terminates in the upper bound. The dotted line shows a less conservative strategy (i.e. thresholds closer to the starting point), requiring less evidence, which leads to the lower bound. The vector represents the drift-rate which is the trend of the accumulation process in that condition block.

Compulsive Disorder (Banca et al., 2015; Erhan & Balcı, 2017; Hauser et al., 2017), Parkinson's Disease (O'Callaghan et al., 2017), Autism Spectrum Disorder (Pirrone et al., 2017), Aphasia (Ratcliff, Perea, et al., 2004; White et al., 2010) and Dyslexia (Zeguers et al., 2011).

1.2. The present study

The fact that the WS cognitive profile exhibits such a multifaceted set of strengths and weaknesses calls for an approach capable of taking different aspects of cognitive processes into account. When participants respond to an experimental (e.g. lexical) trial, their response time is the product of different cognitive processes, some directly relevant to the task (e.g. linguistic difficulty) and some indirectly (e.g. response caution). For this reason, a research methodology is needed that captures such different processes. In the present study, we address this issue by measuring and analysing RT distributions along with performance accuracy. To date, only a few studies have looked at RTs in the context of language and WS (e.g. Karmiloff-Smith et al., 1998; Tyler et al., 1997), or in that of spatial processing (e.g. Karmiloff-Smith et al., 2004; Pani et al., 1999; Scerif et al., 2004). However, studying RTs is crucial for understanding the underlying cognitive processes of an experimental task (Luce, 1986).

To understand which aspects of lexical processing are particularly underdeveloped in WS individuals, we performed a receptive vocabulary task (similar to Clahsen et al., 2004). In this task, participants heard a target auditory stimulus describing a noun and simultaneously they were presented with two pictures on the screen. One picture displayed the noun, whereas the other picture was a distractor displaying another object. The goal of the task was for the participants to match the semantic content that the auditory stimulus elicited with the appropriate picture. The response times and choices were analysed with the DDM to understand how the observable behaviour can be decomposed in cognitive processing components, and which of these components is specifically deviant for WS individuals. Since WS individuals have strong receptive vocabulary ability, we chose a receptive vocabulary task to investigate potential difficulties in receptive vocabulary that are not related to lexical processing per se but rather originate from response caution or perceptual processing (as reflected in the decision threshold and non-decision time parameters of the DDM).

1.2.1. The diffusion decision model

Benefiting from developments in the decision-making literature we used the DDM as a unifying framework for RTs and accuracy. Several studies have applied the DDM to language processing of healthy adults, in particular to lexical decision tasks (e.g. Anders et al., 2015; Anders et al., 2017; Ratcliff, Gomez, et al., 2004; Wagenmakers et al., 2008). The DDM describes how participants gather perceptual information for the response options in a set of trials (e.g. Donkin & Van Maanen, 2014), as well as how much information they need to commit to a decision (a reflection of response caution, e.g. van Maanen et al., 2011; Winkel et al., 2012). In addition, the DDM separates the time of the RT distribution related to information gathering and the time for sensorimotor processes that do not relate to information gathering per se. Taken together, the model is able to decompose the effects of response caution, information processing and sensorimotor processing that are "packaged" in the RT and accuracy data; something that would have been missed if we only look at mean RTs or performance accuracy in isolation.

In our study, we first fit the DDM to the vocabulary task data and then compare WS individuals to a group of 5-year-old healthy preschoolers through the lens of DDM parameters, reflecting individual differences in cognitive processing. Our aim is to investigate whether the latent components of the WS individuals (as measured by the DDM) coincide with those of a group of typically developing 5-year-olds, or whether some aspect of the lexical decision-making process of WS deviates from this group of typical development. Second, we investigate associations between the DDM parameters and intelligence measures. Third, we investigate the association between the DDM parameters and a performance measure on an additional orientation discrimination task. Since WS is characterised with an array of different strengths and weaknesses, we ask the question of whether WS performance in the receptive vocabulary task can predict performance in the orientation discrimination task. Such cross-task association will provide evidence for a task-specific (lexical-only) impairment or a global task-independent impairment in WS.

2. Experimental investigation

2.1. Participants

Three groups participated in the study, namely a WS, a preschooler and a control group (see Table 1 for group descriptives; for participant descriptives see Table 1 in supplementary materials). The WS group consisted of ten adolescents and adults (Mean chronological age = 22 years and 7 months), all previously diagnosed with WS using a fluorescent in situ hybridisation test. The preschooler group consisted of twelve 4–5 year olds of typical development (Mean chronological age = 4 years and 10 months), while the control group included twelve adults (undergraduate students in linguistics; mean chronological age = 22 years and 10 months). All participants were native speakers of Greek. All control participants and all parents or guardians of the preschooler and WS group signed an informed consent.

2.2. Cognitive measures

The screening procedure included two cognitive tests. The first was the *Raven Colour Matrices Test* (Raven, 1998), which measures non-verbal intelligence. The WS and preschooler group did not differ on the basis of their corresponding mental age (MA) according to the Raven test ($t(19.94) = -1.33$, $p = 0.198$). The screening further included the Greek version of the *Peabody Picture Vocabulary Test–Revised* (Simos et al., 2011; PPVT), which measures receptive vocabulary. Although the initial sampling plan dictated three groups of twelve participants, two participants with WS were left out of the study during the screening because of the very poor results on the Raven test.

2.3. Receptive vocabulary task

2.3.1. Materials

The primary task was a receptive vocabulary task, which assessed the participants' receptive vocabulary abilities. The task loosely followed the design of Clahsen et al. (2004) but it was adapted to the binary response structure required for fitting the DDM.

The receptive vocabulary task had two conditions. In one condition, the two presented pictures came from the same semantic category (e.g. an elephant and a lion, both being animals). In the second condition, the two pictures came from distinct semantic categories (e.g. a car and a lion, one being a vehicle and the other being an animal). The conditions were presented in a block design and their order of presentation was counterbalanced across participants.

One hundred and twenty-seven pictures were used from the PD/DPSS dataset of Dell'acqua et al. (2000), 75 of which were used as target words and 52 as distractors (for individual properties of the pictures and distributions across conditions and groups see supplementary materials). The Italian words of the

Table 1. Descriptive statistics of the groups included in the study. "SD" stands for standard deviation.

Group	*n*	Age range (year:month)	Gender proportion (out of total n)	Mean Raven Score (SD)	Mean PPVT Score (SD)
WS	10	13:11–37:01	6 females	14.80 (4.02)	97.70 (38.14)
Preschoolers	12	4:7–5:1	5 females	17.25 (4.61)	80.05 (14.04)
Controls	12	20:04–24:01	6 females	–	–

dataset were translated to Greek. Because the distractor set was smaller than the target set, randomly chosen distractor pictures were reused. The words came from eleven categories provided by the PD/DPSS dataset, namely birds, buildings, clothes, furniture, fruits, instruments, mammals, vegetables, vehicles and weapons, and a mixed one. Word frequencies of the translated target words were retrieved based on the Hellenic National Corpus (Hatzigeorgiu et al., 2000; URL: http://corpus.ilsp.gr/). Word frequencies and number of syllables of the target words were balanced across conditions and groups.

The size of the pictures was kept the same in every trial and both pictures were placed equidistantly from the centre of the screen; one to the left, the other to the right. The correct image was randomly presented either to the left- or to the right-hand side of the screen.

2.3.2. Procedure

Each participant saw 75 trials (half from the same semantic category condition, and half from the different category condition). We opted for this number of trials in order to avoid cognitive overload of WS participants. In every trial a random semantic category was picked and a random word was selected. The algorithm kept track of the presented words, so that target words were shown only once per participant. Before each trial, a fixation cross was presented for one second. The trial was presented until participants provided a response. There was no response deadline. The auditory stimuli were recorded by a native Greek speaker under the same acoustic conditions and all items exhibited the same intonation structure. No word exceeded the number of six syllables.

2.4. Orientation discrimination task

2.4.1. Materials

The secondary task focused on spatial ability and examined the threshold up to which participants were able to recognise degree differences in tilted objects (for a graphical representation of the task see Supplementary Materials). The design of the task followed that of Palomares et al. (2009). Participants were given a Gabor gradient at the upper centre area of the monitor which was the target stimulus. Simultaneously, two Gabor gradients appeared in the bottom left-hand and bottom right-hand side of the screen, both of which differed only in terms of their degree of tilt. The goal of the task was for the participants to decide which of the two provided bottom gradients had the same tilt with the target gradient. One of these had the same orientation with the target gradient and the other was a distractor.

The Gratstim module of Psychopy was used (Peirce, 2007) to generate the Gabor patches. The gradients followed a sinusoidal pattern of black and white shades with a circular Gaussian opacity 3D distribution centred at middle of the gradient. The gradients had sizes of 4 deg of visual angle and were linearly interpolated with texture resolution of 256. The two option gradients were centred at equal distance from the centre of the screen. The tilt direction (whether the target gradient was tilted to the left or right) as well as the position of the correct gradient (whether presented to the left or to the right of the monitor) were randomised on every trial.

The experiment was divided into four conditions, where the target gradient had orientation of −45, 0, 45 and 90 degrees, respectively. The four conditions were presented in blocks and their presentation order was counterbalanced across participants. The number of trials of each condition was determined by the group of examination, a strategy also used in Palomares et al. (2009). In adults, each condition lasted for 35 trails, whereas that number dropped to 25 in preschool children and participants with WS in order to make the task feasible given the attention limitations of these groups.

2.4.2. Procedure

As in Palomares et al. (2009), we used the adaptive staircase method. The experiment started at a 60 deg difference between the choice gradients, and as long as participants were giving a correct response the orientation difference was divided by two in every successive trial. If participants made a mistake, then the orientation difference of the next trial was increased by 20 deg. The same method held for every condition until the condition was finished and thus the degree difference was again set to 60 deg for every new condition block.

Before each trial, a pre-trial fixation cross was presented for one second so that participants could focus their attention on the screen. The duration of trials varied because every trial was terminated by pressing a key. Participants received feedback on their performance in 25% of the trials. The feedback was presented for one second as a recorded message and also as a written message on the screen.

3. Data analysis

For comparative reasons, we first analyse the data of the primary task using standard (yet state-of-the-art) regression analyses. Then, we move on to the cognitive modelling of the component processes using DDM.

The data were preprocessed to preclude outliers from the behavioural analysis. The last 5% of trials per participant were not taken into consideration based on the RT distribution of each participant (Baayen & Milin, 2010). This particular cut-off is considered enough to exclude unreasonably late responses and to prevent the entire elimination of a potential effect that may lie in the right tail of the distribution. Furthermore, responses faster than 250 ms were not taken into consideration, since the tasks did not attempt to elicit fast responses, something that resulted in omitting three responses.

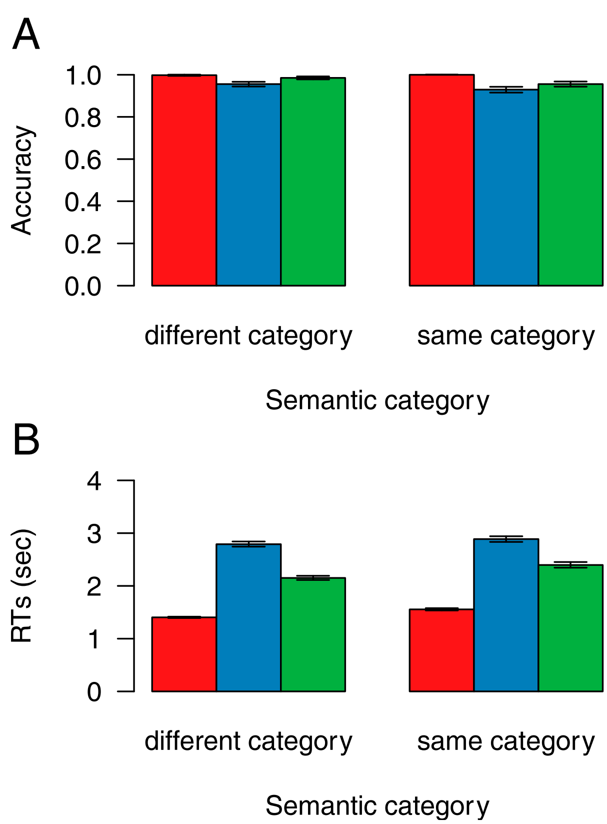

Figure 2. Mean behavioural measures for the receptive vocabulary task across the two conditions. (A) Accuracy; (B) Mean RTs. Errors bars indicate standard errors of the mean. Green bars indicate WS participants, blue indicate preschoolers, while red indicate controls.

For the analysis of the behavioural-descriptive data we used brms (Bürkner, 2017), which implements Bayesian regression modelling in the probabilistic programming language stan (Carpenter et al., 2017). Bayesian statistics offers many advantages over classical null-hypothesis significance testing (e.g. Wagenmakers et al., 2018). We used brms in order to construct mixed-effect (a.k.a. hierarchical) regression models (Baayen et al., 2008; Bates et al., 2015; Bolker et al., 2009; Crawley, 2012).

We modelled the response accuracy of the primary task as a binomial variable. We included the group and the semantic category condition as main effects and, in addition, we modelled their interaction. Furthermore, we included random intercepts per participant as well as per target lexical item, and we also included random slopes for the semantic category and group effects per participant. We used the default prior distributions of brms. We fitted the model for 2000 iterations with 3 independent chains. The first 1000 iterations in each chain were discarded as warmup. The model converged with all $\hat{R} < 1.01$. In the next paragraphs, 95% Highest Density Intervals (HDI) of the coefficients of interest are presented. We interpret a difference between conditions whenever the HDI does not include zero, indicating that the likelihood that a population parameter is not zero exceeds 95%

We replicated the original Clahsen et al. (2004) results for WS even with just two competing stimuli (as compared to 12 or 24 available responses in the original study). Specifically, we found an effect of semantic similarity on receptive vocabulary in the WS group ($95\%_{HDI}$: [−2.56, −0.15]), meaning that lexical items from the same semantic category elicit more errors only for WS individuals (Figure 2(A)). We did not find a semantic similarity effect in the preschooler group ([−1.29, 0.13]) and the control group ([−0.74, 84.75]).[4]

We further evaluated whether there were differences across the three groups. In the different semantic category condition, the control group had higher accuracy than both the WS group ([0.11, 11.28]) and the preschoolers group ([1.77, 12.90]), but the accuracy between the WS group and the preschoolers group did not differ ([−3.07, 0.07]). In the same semantic category conditions, the control group had higher accuracy than both the WS group ([−91.20, −2.34]) and the preschooler group ([−92.10, −3.68]), but, as before, there was no difference in the accuracy of the WS and the preschooler group ([−1.91, 0.37]). To assess whether these no difference conclusions reflect support for the null hypothesis or rather insufficient data to support either the null or the alternative hypothesis, we conducted a Bayes Factor (BF) analysis. BFs indicate how many times likely a hypothesis is over another (cf. Lee & Wagenmakers, 2014). For the preschooler-WS data alone, we used the regression model we fitted above as representing the alternative hypothesis and we further fitted an additional model, representing the null hypothesis, with the only difference that, in the null model, the main effect of the group was removed. Absolute BFs above 3 indicate at least moderate evidence in favour for a hypothesis. The BF in favour of the null hypothesis was 1.53, indicating insufficient data to support either hypothesis. Because of the low sample size within groups (naturally in the context of developmental studies of clinical populations), more data is needed to establish that no difference in mean accuracy between the WS and preschooler groups indicates exactly the same performance accuracy at the level of the WS and preschooler populations.

Furthermore, two interactions were observed. Firstly, we observed an interaction between the control group and the preschooler group across the two conditions ([0.16, 109.81]), and, secondly, we observed an interaction between the control group and the WS across the two conditions ([0.64, 105.19]).

Overall, the semantic similarity condition worked for WS individual only. Also, preschoolers and the WS participants performed in the receptive vocabulary task similarly well, although more data are required to establish that the performance of the two groups is the same.

RT results are shown in Figure 2(B). For the RTs, we modelled the response variable as a shifted log-normal distribution (cf. Wagenmakers & Brown, 2007), which is commonly used because RT distributions have a long right tail and are bound to zero. The same independent variables were included as in the case of response accuracy. We fitted the model with an increased number of iterations (i.e. 4000) in 3 independent chains in order for the model to converge. The first 1000 iterations in each chain were discarded as warmup. The model converged with all $\hat{R} < 1.01$. Posterior predictive checks ensured that the shifted log-normal distribution captured the overall data distribution well.

First, we examined if there was an effect of semantic category condition. The adult group produced slower responses in the same category condition on average as compared to the different category condition ([0.06, 0.28]).[5] However, this effect did not hold for the WS group ([−0.02, 0.23]) and the preschoolers group ([−0.06, 0.16]).

Second, we asked the question whether the three groups differ in their RTs. In the different semantic category condition, mean RTs of the WS group ([0.23, 1.07]) and the preschooler group ([0.62, 1.19]) were slower than those of the control group, but there was no difference between the mean RTs of the WS and the preschooler group ([−0.70, 0.02]). In the same category condition, the WS group ([0.50, 1.09]) and the preschoolers group ([0.15, 1.06]) were again slower than the control group, but there was no difference in the RTs between the preschooler and the WS group ([−0.07, 0.62]). No interactions were observed between the group and the semantic category condition. As with response accuracy, we computed the BF on the preschooler-WS data in order to verify whether the absence of difference in the RT data between these two groups is due to insufficient data or rather support for the null hypothesis (i.e. the two groups have the same mean RT). The BF in favour of the null hypothesis was 1.40 indicating insufficient data to conclude the null or the alternative hypothesis. Therefore, in the case of RTs we found no evidence for a different behavioural pattern between WS individuals and preschoolers.

Overall, we replicated the initial findings of Clahsen et al. (2004). Semantic similarity made WS (but not preschoolers) to produce more errors in receptive vocabulary. As expected, the adult control group exhibited better performance accuracy. The slight differences in the behavioural patterns across groups for accuracy and RTs indicate that the underlying differences in cognitive processing may be multi-faceted, calling for a more fine-grained analysis, which we will now turn too.

3.1. Fitting the diffusion decision model

We fitted a version of the so-called "pure" DDM to the data. This version estimates the basic drift-rate, the threshold, the non-decision time parameters and assumes that they do not vary from trial to trial (the bias parameter of the DDM was set to 0.5 and it was not estimated). We did not include trial-to-trial variability parameters for simplicity because, firstly, the pure DDM version fitted the data well (see below) and, secondly, trial-to-trial variability parameters are sometimes hard to estimate reliably (Boehm et al., 2018).

We fitted the DDM in a Bayesian hierarchical fashion separately per group and experimental condition. We assumed that each individual's parameter estimate comes from a hyper-distribution with the same mean and standard deviation, where the drift-rate, threshold and non-decision time have separate hyper-distributions. Unlike non-hierarchical models, hierarchical models have the advantage of using group level information to constrain individual-level parameters based on the hyper-parameters ("shrinking"), something that leads to more reliable parameter estimates (Farrell & Lewandowsky, 2018; Farrell & Ludwig, 2008). Because we are interested in making comparisons not only across groups but also across conditions, we estimated two different hyper mean and standard deviation parameters for the drift-rate, threshold, decision-making (each for either of two conditions: words from the same semantic category, words from the different semantic category).

We fitted a Hierarchical Bayesian DDM using a non-centred parametrization in the stan programming language (Carpenter et al., 2017). The priors were the standard normal distribution for all parameters. Participant-level parameters were rescaled from an unbounded scale to a zero-bounded one. According to standard practice, we exponentiated the drift-rate and threshold (to be bound from 0 to infinity), and we applied the CDF of the standard normal distribution to the non-decision scaled by the maximum observed RT. We adapted code from the *hBayesDM* package (Ahn et al., 2017). We run three independent chains for 10000 iterations, the first 1000 of which was used for warmup. The diffusion coefficient was constrained to 1. RTs were fitted in seconds. All models converged with $\hat{R}$ lower than 1.01 with effective sample size higher than 400 in all parameters.

The fit of the DDM to the data of the linguistic task are shown in Figure 3 for five RT quantiles and performance accuracy of correct responses. Predictions of the DDM were sampled based on mean posterior parameters of each participant. The model fits both RTs and accuracy very well. The notable deviation from this pattern is in the incorrect RT distribution; since patterns made few errors, these distributions are often hard to predict and are not shown here.

3.2. Mean posterior parameters

After fitting the DDM to the data we looked at the distributions of the DDM's posterior parameter distributions. The mean hierarchical posterior densities of all parameters of the receptive vocabulary task are shown in Figure 4. The shaded areas under the densities show 95% HDIs. The control group was characterised by high drift-rates (same category $95\%_{\mathrm{HDI}}$: [0.898, 1.422]; different category: [0.865, 1.276]) that were higher than the WS group (same category: [−0.167, 0.574]; different category: [0.164, 0.829]) and the preschooler group (same category: [−0.321, 0.082]; different category: [−0.188, 0.326]) in both conditions. This implies that the speed of lexical accumulation for the control group was faster than the other groups, but, interestingly, the WS and preschooler groups showed similar accumulation speeds.

In addition, the control group showed an elevated threshold when the words came from the same semantic category ([1.304, 1.938]) as compared to the other condition ([0.893, 1.280]). WS and preschooler participant had similar threshold levels as the control group in the different category condition ([1.011, 1.327] and [1.089, 1.329], respectively), but lower threshold than the control group in the same category condition ([0.945, 1.277], [0.972, 1.215]). This indicates that WS and preschooler participants exhibited the same levels of response caution irrespectively of the condition, but the control group was more cautious in the same category condition.

Like their threshold, the control group's non-decision time was also affected by the semantic category condition; non-decision time was shorter in the same category condition ([0.073, 0.818]) but larger in the different category condition ([0.822, 1.219]). The WS and preschooler group did not differ in their NDT either across conditions or between them (WS, same category: [0.725, 1.097], different category: [0.710, 1.156]; preschoolers, same category [0.863, 1.161], different category [0.750, 1.063]). In addition, the control group's mean non-decision time was the same like the other two groups in the different category condition, but it was shorter only as compared to that of the preschoolers in the same category condition.

Overall, the hierarchical mean drift-rate, decision threshold and non-decision time posterior distributions of WS and preschoolers did not differ in any condition. In addition, the control group had higher drift-rate than both other groups generally.

3.3. Drift-rate and intelligence measures

Since the DDM characterises the data in terms of latent model parameters, we are able to regress other variables to the model's parameters of the lexical data and not simply on the data themselves. In particular, we evaluated whether the rate of lexical accumulation (drift-rate) of the DDM is affected by either verbal or non-verbal intelligence, as measured by the PPVT and the Raven. Because the drift-rate reflects the difficulty of the task, we hypothesised that verbal intelligence measures would correlate with the drift-rate of the lexical task.

To verify this hypothesis, we reran the model with the following slight modification: we used the standardised Raven and PPVT scores of participants as linear predictors of the drift-rate. The overall drift-rate, therefore, took the form of $\delta = \delta_0 + \beta_{\mathrm{raven}} * \mathrm{RAVEN} + \beta_{\mathrm{PPVT}} * \mathrm{PPVT}$. If the PPVT or/and the Raven score explain more variance in the drift-rate than the intercept (δ_0), then the HDIs of the coefficients β_{raven} and β_{PPVT} should not include zero. The coefficients had standard normal

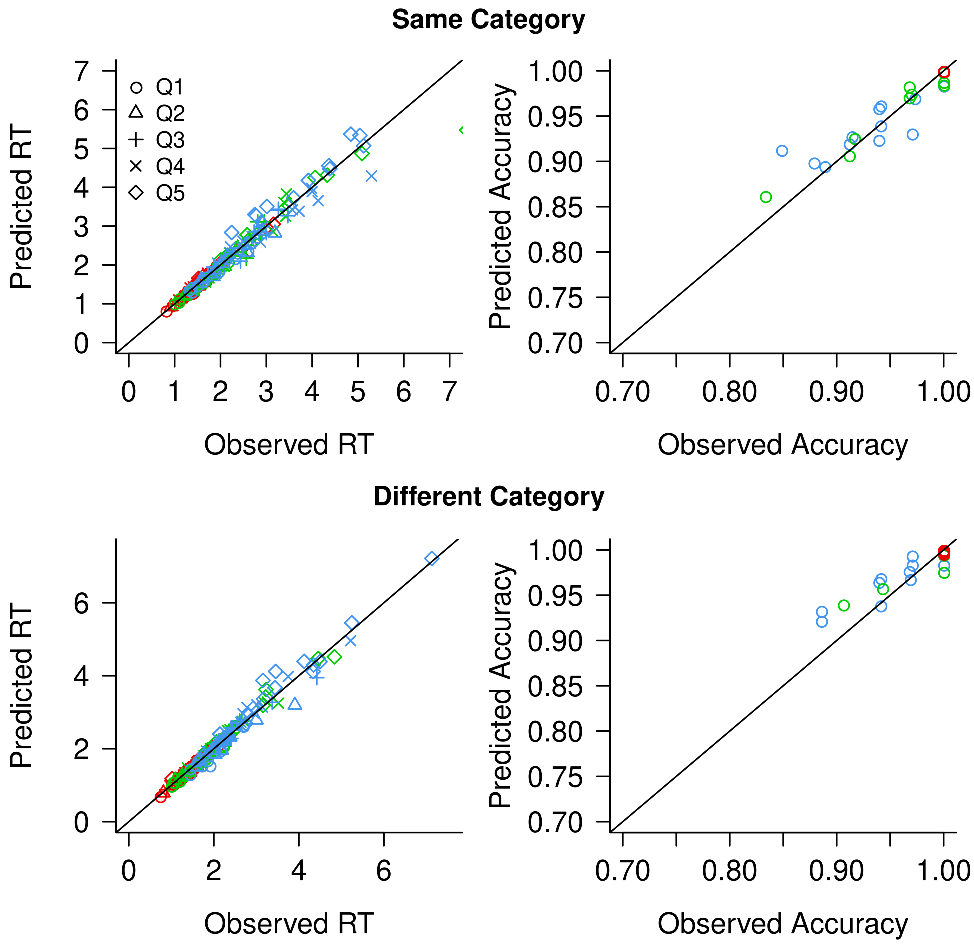

Figure 3. Fit of the DDM to data of the receptive vocabulary task. Observed vs. predicted accuracy and five RT quantiles (10% (Q1), 30% (Q3), 50% (Q5), 70% (Q7) and 90% (Q9)) are shown per participants and conditions (i.e. target options coming from similar or dissimilar semantic category) only for correct responses, as well as response accuracy. Green characters indicate WS participants, blue indicate preschoolers, while red indicate controls. Lines represent the identity function.

priors and they were not modelled hierarchically (since each participant had only one Raven and PPVT score, respectively).

Table 2 presents the posterior results of the coefficients. As hypothesised, we see that the Raven task does not have an effect on the drift-rate. On the other hand, the PPVT, as a measure of verbal intelligence, is related to the drift-rate of the WS group. This means that WS individuals with higher PPVT scores accumulate lexical information faster on average. On the other hand, no association between the PPVT scores and drift-rates was observed in the preschooler group.

3.4. Associations in WS performance across tasks

Relating the cognitive profiles of the participants with performance in the orientation discrimination task can reveal similar ways in which participants responded to the tasks. In particular, if the same participant found both tasks equally difficult or easy, this could be evidence for common cognitive strategies that affect both tasks simultaneously. It would be therefore evidence for across-domain strengths (or impairments) irrespective of the specific cognitive domain at which the linguistic or the orientation discrimination task aim. Although not the focus of the current study, for completeness we report behavioural-descriptive analyses of the orientation discrimination task in Supplementary Materials.

To relate the performance in both tasks, we looked at the correlation between the mean posterior drift-rates (averaged across conditions) in the lexical task and the mean orientation thresholds (averaged across conditions) in the orientation discrimination task. The posterior drift-rates from the lexical task appeared to negatively correlate with the orientation discrimination threshold. Nevertheless, when applying a multiple regression model, we found that this is a spurious association since the group to which participants belong explains almost all variance in lexical drift-rates (R^2 =

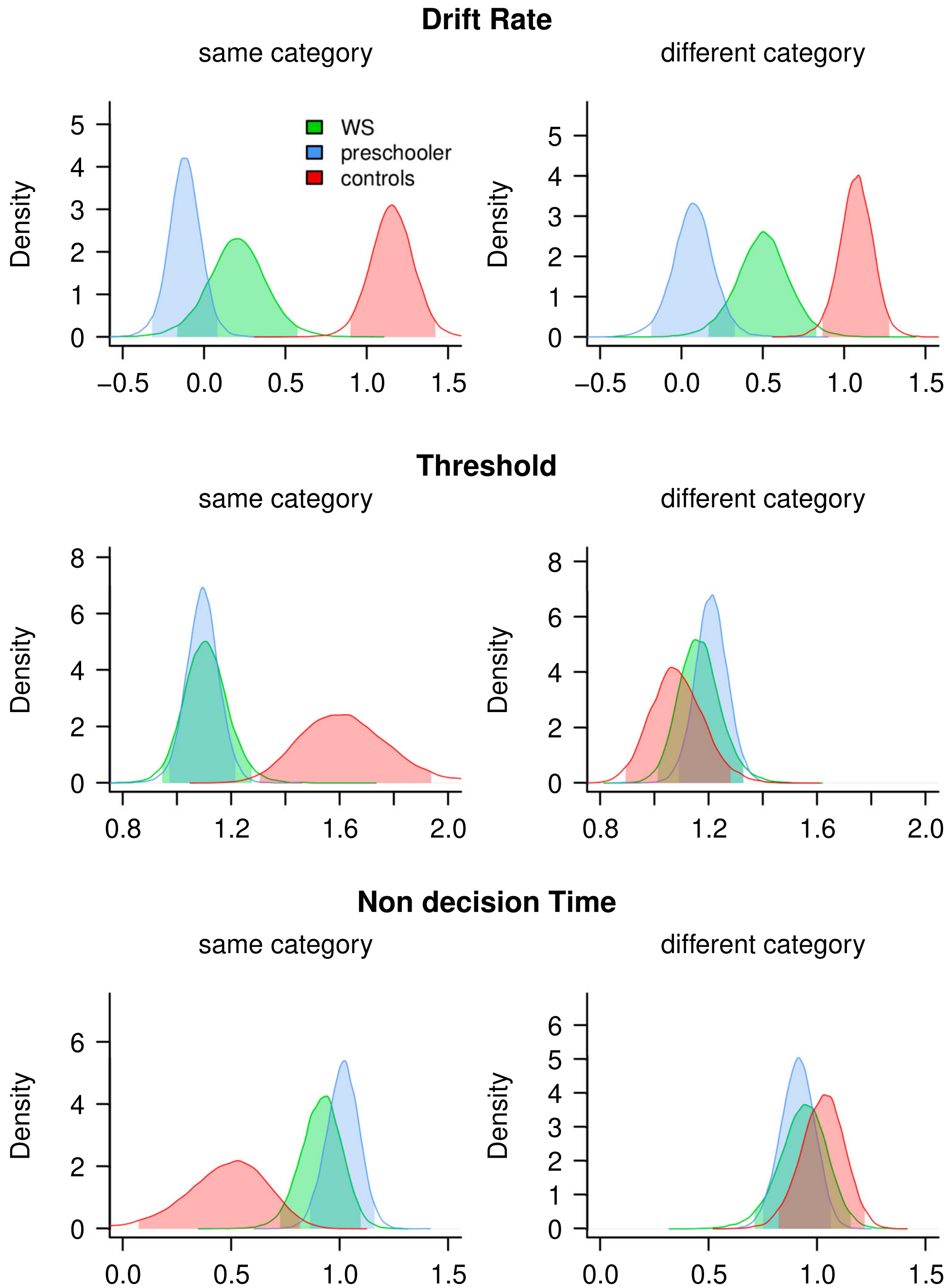

Figure 4. DDM posterior densities of the hierarchical mean parameters of drift-rate, threshold and non-decision time in the receptive vocabulary task. Shaded areas contain the 95% HDI. Densities are given per group (green indicate WS participants, blue indicate preschoolers, while red indicate controls) and condition (left column: semantically unrelated distractor; right column: semantically related distractor).

Table 2. HDIs of the Raven and PPVT regressors on the drift-rate of the lexical task. The results are given per group and condition. HDIs in bold are the ones that do not include zero.

Group	Condition	HDI Raven	HDI PPVT
WS	Same category	[−0.388, 0.378]	**[0.062, 0.881]**
	Different category	[−0.587, 0.332]	**[0.115, 1.086]**
Preschoolers	Same category	[−0.162, 0.144]	[−0.093, 0.304]
	Different category	[−0.230, 0.224]	[−0.126, 0.423]

72%), while the orientation threshold explains very little variance ($R^2 < 0.01\%$). This means that the group is the most important factor since the group factor alone explains well enough that the control group has higher lexical drift-rates and lower orientation thresholds (and vice versa for the rest of the groups). Therefore, our data are insufficient to provide evidence for an association between lexical drift-rates and orientation thresholds at the individual level.

4. Discussion

In this paper, we applied the popular DDM to people with WS in order to investigate their cognitive characteristics when a receptive vocabulary (lexical) task is administered to them. The WS individuals were compared to a group of typically developed 5-year-old preschoolers with comparable non-verbal intelligence, while a control group of typically developed adults was also included. According to the parameters of the DDM, we found no evidence for a difference between WS and preschooler participants in all latent components of the model in the lexical task. Specifically, the way WS and preschooler participants accumulated information in order to respond (i.e. the drift-rate), their response caution capacity (i.e. the decision threshold) and their sensorimotor processing (i.e. non-decision time) were all found to be similar across experimental conditions. Therefore, we conclude that, according to the DDM, the performance of WS is similar to their 5-year-old counterparts of typical development.

We successfully replicated some of the results of Clahsen et al. (2004): WS individuals find it more difficult to correctly identify lexical items between two semantically related words. However, this effect was found only when looking at the performance accuracy data alone. Also, no difference was found between the WS and the preschooler group in either performance accuracy or RTs. According to the posterior parameters of the DDM, the speed of lexical accumulation (as reflected in their drift-rate) in WS overlapped for different semantic relatedness conditions. Therefore, when RTs and accuracy are taken together in the context of the DDM, we did not find enough evidence for an effect of semantic relatedness in WS as well as between the WS and the preschooler group. This conclusion is in contrast to Clahsen et al. (2004), who found that WS performed worse than a control group in semantically related items. Taken together, our results do not support the view (e.g. found in Bellugi et al., 1990; Jarrold et al., 1999; Vicari et al., 2004) that the receptive vocabulary of WS individuals is developed more than one would expect on the basis of their non-verbal mental age. Our WS group and the preschooler group of comparable non-verbal mental age showed similar lexical accumulation rates, which implies that they found the task equally difficult in terms of lexical demands. Nevertheless, more data are needed to further corroborate this finding.

In addition, the DDM analysis did not show any difference in response caution (as reflected in the decision threshold) between the WS and the preschooler group. However, previous studies have suggested that WS exhibits inhibition deficits (i.e. difficulty in withholding a response; e.g. Greer et al., 2013), which may be caused by reduced frontostriatial activity (e.g. Mobbs et al., 2007). There is evidence that the decision threshold of the DDM correlates with changes in activation in striatal and pre-supplementary motor areas in order to optimise behaviour for different environmental demands (e.g. Forstmann et al., 2008; Ivanoff et al., 2008; van Maanen et al., 2011; van Maanen et al., 2016). However, the fact that, in our study, the decision threshold was observed to be similar between WS and preschoolers does not necessarily imply that response caution is not impaired in WS. Our receptive vocabulary task might not have induced such increased response caution demands that would show a difference between the two groups. Future research may explicitly evaluate lexical decisions under different behaviour optimisation demands (such as the speed-accuracy trade-off, cf. Mulder et al., 2010; or response inhibition) in order to better understand WS response caution during lexical processing.

Our results showed that the speed of lexical accumulation of WS individuals (as reflected in the drift-rate of the DDM) is associated with their receptive vocabulary intelligence scores but not with those of non-verbal intelligence. This is, firstly, in line with previous literature showing that the drift-rate is associated with intelligence measures (Ratcliff et al., 2008; van Ravenzwaaij et al., 2011), but it also validates that the DDM's measure of lexical accumulation directly reflects verbal intelligence.

Furthermore, we observed that the drift-rates of the receptive vocabulary task and the orientation discrimination threshold measured in the secondary task were associated, namely those participants that found it more difficult to accumulate lexical information also exhibited poorer orientation discrimination capacity. However, we found that this association was mediated by the group individuals belonged to (e.g. the control group had higher lexical drift-rates and lower orientation thresholds overall) and after controlling for the effect of group, there was not enough evidence to support a direct association between orientation discrimination threshold and lexical drift-rates. Since the lexical ability of WS individuals cannot be directly predicted by the performance of the WS individuals in the perceptual task, it implies that lexical processing in WS has domain-specific characteristics. In addition, the observation that these accumulation rates did not correlate with scores on a test of non-verbal intelligence (Raven) points in the same direction, suggesting that impairment to the lexical access system is not associated with a lower intelligence per se. Nevertheless, these results should be interpreted with caution given the increased sample size required for correlations to capture small or medium associations. Future research may resort to joint DDM modelling of lexical and perceptual performance and estimate the association of lexical and perceptual drift-rates parameters of the DDM (e.g. Palestro et al., 2018).

Taken together, our results support an approach to WS development that characterises it as a diverse set of strengths and weaknesses by distinguishing domain-specific (i.e. task-specific) and domain-general (i.e. task-independent) cognitive impairments (Landau & Ferrara, 2013). WS receptive vocabulary ability, although chronologically delayed, matches the developmental stage of typically developing 5-year-old children as described by all latent components of the DDM. This is in line with the view that the development of WS is characterised by a normal but slowed developmental trajectory (for a review see Landau & Ferrara, 2013). Nevertheless, more cross-sectional research of the developmental trajectory of WS lexical abilities is required to further establish this view.

It is worth-noting that in this paper we used the DDM as a measurement tool and not as theory for WS decision-making (on measurement tool vs. theory distinction see Evans & Wagenmakers, 2019). We do not propose that the DDM is the best theoretical explanation of how WS individuals make lexical or perceptual decisions. Our approach was to use the latent components of the DDM as a tool to map the behavioural data of WS participants to the models' parameters in order to achieve a higher-order cognitive characterisation of WS behaviour. Future research might evaluate the explanatory power of the DDM in comparison to competing accounts of WS decision-making. Nevertheless, we recommend that researchers use cognitive models in WS research in order to obtain a more informative picture of the cognitive profile of WS individuals beyond simply observing behavioural measures (Anders et al., 2016; Donkin et al., 2011; Voss et al., 2013).

4.1. Limitations of the study

The present paper used the DDM as a cognitive framework to the analysis of lexical aspects of cognitive processing in people with WS. The adopted framework illustrated the way task-specific information processing, response caution and non-decision time can provide useful insights on the way WS process linguistic stimuli. Future modelling approaches may benefit from integrating these results along two dimensions: that of the breadth of cognitive aspects of WS taken into account, and that of developmental trajectory. Specifically, apart from verbal and non-verbal intelligence, future modelling approaches may aim to explain WS executive functions (e.g. working memory, attention or motor planning). Such investigations will enrich the data-model correlations possibly attributing different standardised measures to different model components (Turner et al., 2017; van Maanen & Miletić, 2020). In addition, the investigation of WS developmental trajectory would be richer if future studies included not only a mental-age-matched group and a control group but several ones that are chronologically in-between. The inclusion of such intermediate groups will provide a fuller developmental trajectory of the phenomena under investigation, which may be insightful of the nature of critical slow-down in WS development (Landau & Ferrara, 2013). In addition, it is worth mentioning that since our tasks did not assess WS performance across levels of linguistic (e.g. phonology in comparison with semantics) and visuo-spatial analysis (e.g. orientation integration in comparison with orientation discrimination), our results are not informative about the debate regarding imbalances within the various WS cognitive domains (e.g. Camp & Karmiloff-Smith, 2015; Karmiloff-Smith, 1998; Laing et al., 2001; Tyler et al., 1997).

5. Conclusions

The current study applied the DDM to data of WS in a receptive vocabulary task; thus, being one of the few cases of cognitive modelling applied to WS data up to date. Based on all latent components of the DDM, WS participants had similar performance to that of healthy 5-year-olds with a comparable mental age. We believe future research can benefit from the use and comparison of many groups from different phases of typical development as well as from the thorough application of additional information processing cognitive models that are able to distinguish different aspects of processing entrenched in the WS behavioural patterns.

Notes

1. Though several theoretical frameworks have been proposed to explaining the complex array of behavioural, cognitive and neurological characteristics of individuals with WS (e.g., nativist approaches (Bellugi et al., 1988; Pinker, 1999), neuroconstructivist ones (Karmiloff-Smith, 1998, 2009; Karmiloff-Smith et al., 2003), hypotheses of dorsal-stream disorder (Atkinson et al., 2006; Landau & Hoffman, 2005), among others), a universally

accepted framework still remains elusive. An extensive discussion is beyond the scope of the current paper.
2. Some versions of the DDM model across-trial variability in the non-decision time parameter resulting in a distribution (and not point estimates) of the non-decision time. However, in our analysis we assumed that across-trial variability is zero for simplicity.
3. In addition, the DDM employs several parameters that encode the extent of parameter variability from trial to trial (Ratcliff, 1978; Ratcliff & Rouder, 1998). Parameter variability has been primarily used to make the model fit human RT data better.
4. HDIs of coefficients of response accuracy are given on the logit scale.
5. HDIs of coefficients of RTs are given in log(RT) scale.

Acknowledgements

We would like to thank all children and their families for participating in the study, as well as the "Aghia Sofia" General Children's Hospital and Dr. Eleni Frissira for granting us access to the WS participants. We would like to thank Ioanna Kolokytha for helping us with data collection. Also, thanks to Kevin Trutmann for his helpful comments on an earlier draft of the paper. Parts of the paper were presented in the 9th Athens Postgraduate and PhD Candidate conference in October 2017 in Greece. We also acknowledge the Onassis Foundation for funding the first author through its Scholarship programme for Hellenes. All errors are our own.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Alexander S. Onassis Public Benefit Foundation.

ORCID

Dimitris Katsimpokis http://orcid.org/0000-0002-4073-0206
Leendert van Maanen http://orcid.org/0000-0001-9120-1075
Spyridoula Varlokosta http://orcid.org/0000-0002-9479-7887

References

Ahn, W.-Y., Haines, N., & Zhang, L. (2017). Revealing neurocomputational mechanisms of reinforcement learning and decision-making with the hBayesDM package. *Computational Psychiatry*, *1*(0), 24–57. https://doi.org/10.1162/CPSY_a_00002

Anders, R., Alario, F.-X., & van Maanen, L. (2016). The shifted wald distribution for response time data analysis. *Psychological Methods*, *21*(3), 309–327. https://doi.org/10.1037/met0000066

Anders, R., Riès, S., van Maanen, L., & Alario, F.-X. (2015). Evidence accumulation as a model for lexical selection. *Cognitive Psychology*, *82*, 57–73. https://doi.org/10.1016/j.cogpsych.2015.07.002

Anders, R., Riès, S., van Maanen, L., & Alario, F.-X. (2017). Lesions to the left lateral prefrontal cortex impair decision threshold adjustment for lexical selection. *Cognitive Neuropsychology*, *34*(1–2), 1–20. https://doi.org/10.1080/02643294.2017.1282447

Ansari, D., Donlan, C., & Karmiloff-Smith, A. (2007). Typical and atypical development of visual estimation abilities. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *43*(6), 758–768. https://doi.org/10.1016/S0010-9452(08)70504-5

Archambeau, K., Forstmann, B. U., van Maanen, L., & Gevers, W. (2020). Proactive interference in aging: A model-based study. *Psychonomic Bulletin & Review*, *27*(1), 130–138. https://doi.org/10.3758/s13423-019-01671-0

Atkinson, J., Braddick, O., Rose, F. E., Searcy, Y. M., Wattam-Bell, J., & Bellugi, U. (2006). Dorsal-stream motion processing deficits persist into adulthood in Williams syndrome. *Neuropsychologia*, *44*(5), 828–833. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2005.08.002

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28. https://doi.org/10.21500/20112084.807

Banca, P., Vestergaard, M. D., Rankov, V., Baek, K., Mitchell, S., Lapa, T., Castelo-Branco, M., & Voon, V. (2015). Evidence accumulation in obsessive-compulsive disorder: The role of uncertainty and monetary reward on perceptual decision-making thresholds. *Neuropsychopharmacology*, *40*(5), 1192–1202. https://doi.org/10.1038/npp.2014.303

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

Bellugi, U., Bihrle, A., Jernigan, T., Trauner, D., & Doherty, S. (1990). Neuropsychological, neurological, and neuroanatomical profile of Williams syndrome. *American Journal of Medical Genetics. Supplement*, *6*, 115–125. https://doi.org/10.1002/ajmg.1320370621

Bellugi, U., Lichtenberger, L., Jones, W., Lai, Z., & St George, M. (2000). I. The neurocognitive profile of Williams syndrome: A complex pattern of strengths and weaknesses. *Journal of Cognitive Neuroscience*, *12*(Suppl 1), 7–29. https://doi.org/10.1162/089892900561959

Bellugi, U., Marks, A., Bihrle, A., & Sabo, H. (1988). Dissociation between language and cognitive functions in Williams syndrome. In D. Bishop & K. Mogford (Eds.), *Language development in exceptional cirvumstances* (pp. 177–189). Churchill Livingstone.

Bennett, F. C., LaVeck, B., & Sells, C. J. (1978). The Williams Elfin Facies syndrome: The psychological profile as an aid in syndrome identification. *Pediatrics*, *61*(2), 372–373. https://doi.org/10.1542/peds.65.2.372

Boehm, U., Annis, J., Frank, M. J., Hawkins, G. E., Heathcote, A., Kellen, D., Krypotos, A.-M., Lerche, V., Logan, G. D., Palmeri, T. J., van Ravenzwaaij, D., Servant, M., Singmann, H., Starns, J. J., Voss, A., Wiecki, T. V., Matzke, D., & Wagenmakers, E.-J. (2018). Estimating across-trial variability parameters of the diffusion decision model: Expert advice and recommendations. *Journal of Mathematical Psychology*, *87*, 46–75. https://doi.org/10.1016/j.jmp.2018.09.004

Bogacz, R., Wagenmakers, E. J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed–accuracy tradeoff. *Trends in Neurosciences*, *33*(1), 10–16. https://doi.org/10.1016/j.tins.2009.09.002

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, *24*(3), 127–135. https://doi.org/10.1016/j.tree.2008.10.008

Brock, J. (2007). Language abilities in Williams syndrome: A critical review. *Development and Psychopathology*, *19*(01), 97–127. https://doi.org/10.1017/S095457940707006X

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*(3), 153–178. https://doi.org/10.1016/j.cogpsych.2007.12.002

Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Camp, J. S., & Karmiloff-Smith, A. (2015). Williams syndrome. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (pp. 579–583). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.55055-7

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Carrasco, X., Castillo, S., Aravena, T., Rothhammer, P., & Aboitiz, F. (2005). Williams syndrome: Pediatric, neurologic, and cognitive development. *Pediatric Neurology*, *32*(3), 166–172. https://doi.org/10.1016/j.pediatrneurol.2004.09.013

Clahsen, H., & Almazan, M. (1998). Syntax and morphology in Williams syndrome. *Cognition*, *68*(3), 167–198. https://doi.org/10.1016/s0010-0277(98)00049-3

Clahsen, H., & Temple, C. (2004). Lexical and morphological skills in English-speaking children with Williams syndrome. In S. Bartke & J. Siegmueller (Eds.), *Williams syndrome across languages* (pp. 221–244). https://doi.org/10.1075/lald.36.14cla

Crawley, M. J. (2012). Mixed-effects models. In *The R book* (pp. 681–714). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118448908.ch19

Dell'acqua, R., Lotto, L., & Job, R. (2000). Naming times and standardized norms for the Italian PD/DPSS set of 266 pictures: Direct comparisons with American, English, French, and Spanish published databases. *Behavior Research Methods, Instruments, & Computers*, *32*(4), 588–615. https://doi.org/10.3758/BF03200832

Donkin, C., Brown, S., & Heathcote, A. (2011). Drawing conclusions from choice response time models: A tutorial using the linear ballistic accumulator. *Journal of Mathematical Psychology*, *55*(2), 140–151. https://doi.org/10.1016/j.jmp.2010.10.001

Donkin, Chris, & Van Maanen, Leendert. (2014). Piéron's Law is not just an artifact of the response mechanism. *Journal of Mathematical Psychology*, *62-63*, 22–32. https://doi.org/10.1016/j.jmp.2014.09.006

Elliott, D., Welsh, T. N., Lyons, J., Hansen, S., & Wu, M. (2006). The visual regulation of goal-directed reaching movements in adults with Williams syndrome, Down syndrome, and other developmental delays. *Motor Control*, *10*(1), 34–54. https://doi.org/10.1123/mcj.10.1.34

Erhan, C., & Balcı, F. (2017). Obsessive compulsive features predict cautious decision strategies. *Quarterly Journal of Experimental Psychology*, *70*(1), 179–190. https://doi.org/10.1080/17470218.2015.1130070

Evans, N. J., & Wagenmakers, E.-J. (2019). *Evidence accumulation models: Current limitations and future directions*. https://doi.org/10.31234/osf.io/74df9

Farrell, S., & Lewandowsky, S. (2018). *Computational modeling of cognition and behavior*. Cambridge University Press.

Farrell, S., & Ludwig, C. J. H. (2008). Bayesian and maximum likelihood estimation of hierarchical response time models. *Psychonomic Bulletin & Review*, *15*(6), 1209–1217. https://doi.org/10.3758/PBR.15.6.1209

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., von Cramon, D. Y., Ridderinkhof, K. R., & Wagenmakers, E.-J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences*, *105*(45), 17538–17542. https://doi.org/10.1073/pnas.0805903105

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E.-J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual Review of Psychology*, *67*(1), 641–666. https://doi.org/10.1146/annurev-psych-122414-033645

Grant, J., Valian, V., & Karmiloff-Smith, A. (2002). A study of relative clauses in Williams syndrome. *Journal of Child Language*, *29*(2), 403–416. https://doi.org/10.1017/S030500090200510X

Greer, J., Riby, D. M., Hamiliton, C., & Riby, L. M. (2013). Attentional lapse and inhibition control in adults with Williams syndrome. *Research in Developmental Disabilities*, *34*(11), 4170–4177. https://doi.org/10.1016/j.ridd.2013.08.041

Hatzigeorgiu, N., Gavrilidou, M., Piperidis, S., Carayannis, G., Papakostopoulou, A., Spiliotopoulou, A., Vacalopoulou, A., Labropoulou, P., Mantzari, E., Papageorgiou, H., & Demiros, I. (2000). Design and implementation of the online ILSP Greek Corpus. In *Proceedings of the LREC 2000 Conferences* (pp. 1737–1742).

Hauser, T. U., Allen, M., Rees, G., & Dolan, R. J. (2017). Metacognitive impairments extend perceptual decision making weaknesses in compulsivity. *Scientific Reports*, *7*(1), 6614. https://doi.org/10.1038/s41598-017-06116-z

Hudson, K. D., & Farran, E. K. (2011). Executive function and motor planning. In E. K. Farran & A. Karmiloff-Smith (Eds.), *Neurodevelopmental disorders across the lifespan: A neuroconstructivist approach* (pp. 165–181). OUP.

Ivanoff, J., Branning, P., Marois, R., & Pelli, D. G. (2008). FMRI evidence for a dual process account of the speed-accuracy tradeoff in decision-making. *Plos One*, *3*(7), e2635. https://doi.org/10.1371/journal.pone.0002635

Jarrold, C., Baddeley, A. D., & Hewes, A. K. (1999). Genetically dissociated components of working memory: Evidence from down's and Williams syndrome. *Neuropsychologia*, *37*(6), 637–651. https://doi.org/10.1016/s0028-3932(98)00128-6

Joffe, V., & Varlokosta, S. (2007). Language abilities in Williams syndrome: Exploring comprehension, production and repetition skills. *Advances in Speech Language Pathology*, *9* (3), 213–225. https://doi.org/10.1080/14417040701261517

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences*, *2*(10), 389–398. https://doi.org/10.1016/S1364-6613(98)01230-3

Karmiloff-Smith, A. (2009). Nativism versus neuroconstructivism: Rethinking the study of developmental disorders. *Developmental Psychology*, *45*(1), 56–63. https://doi.org/10.1037/a0014506

Karmiloff-Smith, A., Brown, J. H., Grice, S., & Paterson, S. (2003). Dethroning the myth: Cognitive dissociations and innate modularity in Williams syndrome. *Developmental Neuropsychology*, *23*(1–2), 227–242. https://doi.org/10.1080/87565641.2003.9651893

Karmiloff-Smith, A., Grant, J., Berthoud, I., Davies, M., Howlin, P., & Udwin, O. (1997). Language and Williams syndrome: How intact is 'intact'? *Child Development*, *68*(2), 246–262. https://doi.org/10.1111/j.1467-8624.1997.tb01938.x

Karmiloff-Smith, A., Thomas, M., Annaz, D., Humphreys, K., Ewing, S., Brace, N., Duuren, M., Pike, G., Grice, S., & Campbell, R. (2004). Exploring the Williams syndrome face-processing debate: The importance of building developmental trajectories. *Journal of Child Psychology and Psychiatry*, *45*(7), 1258–1274. https://doi.org/10.1111/j.1469-7610.2004.00322.x

Karmiloff-Smith, A., Tyler, L. K., Voice, K., Sims, K., Udwin, O., Howlin, P., & Davies, M. (1998). Linguistic dissociations in Williams syndrome: Evaluating receptive syntax in on-line and off-line tasks. *Neuropsychologia*, *36*(4), 343–351. https://doi.org/10.1016/S0028-3932(97)00133-4

Krishnan, S., Bergström, L., Alcock, K. J., Dick, F., & Karmiloff-Smith, A. (2015). Williams syndrome: A surprising deficit in oromotor praxis in a population with proficient language production. *Neuropsychologia*, *67*, 82–90. https://doi.org/10.1016/j.neuropsychologia.2014.11.032

Laing, E., Hulme, C., Grant, J., & Karmiloff-Smith, A. (2001). Learning to read in Williams syndrome: Looking beneath the surface of atypical reading development. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *42* (6), 729–739. https://doi.org/10.1111/1469-7610.00769

Laing, E., & Jarrold, C. (2007). Comprehension of spatial language in Williams syndrome: Evidence for impaired spatial representation of verbal descriptions. *Clinical Linguistics & Phonetics*, *21*(9), 689–704. https://doi.org/10.1080/02699200701541441

Lakusta, L., Dessalegn, B., & Landau, B. (2010). Impaired geometric reorientation caused by genetic defect. *Proceedings of the National Academy of Sciences*, *107*(7), 2813–2817. https://doi.org/10.1073/pnas.0909155107

Landau, B., & Ferrara, K. (2013). Space and language in Williams syndrome: Insights from typical development. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(6), 693–706. https://doi.org/10.1002/wcs.1258

Landau, B., & Hoffman, J. E. (2005). Parallels between spatial cognition and spatial language: Evidence from Williams syndrome. *Journal of Memory and Language*, *53*(2), 163–185. https://doi.org/10.1016/j.jml.2004.05.007

Landau, B., Hoffman, J. E., & Kurz, N. (2006). Object recognition with severe spatial deficits in Williams syndrome: Sparing and breakdown. *Cognition*, *100*(3), 483–510. https://doi.org/10.1016/j.cognition.2005.06.005

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Libertus, M. E., Feigenson, L., Halberda, J., & Landau, B. (2014). Understanding the mapping between numerical approximation and number words: Evidence from Williams syndrome and typical development. *Developmental Science*, *17*(6), 905–919. https://doi.org/10.1111/desc.12154

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. Oxford University Press.

Martens, M. A., Wilson, S. J., & Reutens, D. C. (2008). Research review: Williams syndrome: A critical review of the cognitive, behavioral, and neuroanatomical phenotype. *Journal of Child Psychology and Psychiatry*, *49*(6), 576–608. https://doi.org/10.1111/j.1469-7610.2008.01887.x

Miletić, S., & van Maanen, L. (2019). Caution in decision-making under time pressure is mediated by timing ability. *Cognitive Psychology*, *110*, 16–29. https://doi.org/10.1016/j.cogpsych.2019.01.002

Mobbs, D., Eckert, M. A., Mills, D., Korenberg, J., Bellugi, U., Galaburda, A. M., & Reiss, A. L. (2007). Frontostriatal dysfunction during response inhibition in Williams syndrome. *Biological Psychiatry*, *62*(3), 256–261. https://doi.org/10.1016/j.biopsych.2006.05.041

Mulder, M. J., Bos, D., Weusten, J. M. H., van Belle, J., van Dijk, S. C., Simen, P., van Engeland, H., & Durston, S. (2010). Basic impairments in regulating the speed-accuracy tradeoff predict symptoms of attention-deficit/hyperactivity disorder. *Biological Psychiatry*, *68*(12), 1114–1119. https://doi.org/10.1016/j.biopsych.2010.07.031

O'Callaghan, C., Hall, J. M., Tomassini, A., Muller, A. J., Walpola, I. C., Moustafa, A. A., Shine, J. M., & Lewis, S. J. G. (2017). Visual hallucinations are characterized by impaired sensory evidence accumulation: Insights from hierarchical drift diffusion modeling in Parkinson's disease. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *2*(8), 680–688. https://doi.org/10.1016/j.bpsc.2017.04.007

Palestro, J. J., Bahg, G., Sederberg, P. B., Lu, Z.-L., Steyvers, M., & Turner, B. M. (2018). A tutorial on joint models of neural and behavioral measures of cognition. *Journal of Mathematical Psychology*, *84*, 20–48. https://doi.org/10.1016/j.jmp.2018.03.003

Palomares, M., Landau, B., & Egeth, H. (2009). Orientation perception in Williams syndrome: Discrimination and integration. *Brain and Cognition*, *70*(1), 21–30. https://doi.org/10.1016/j.bandc.2008.11.007

Pani, J. R., Mervis, C. B., & Robinson, B. F. (1999). Global spatial organization by individuals with Williams syndrome. *Psychological Science*, *10*(5), 453–458. https://doi.org/10.1111/1467-9280.00186

Pedersen, M. L., Frank, M. J., & Biele, G. (2017). The drift diffusion model as the choice rule in reinforcement learning. *Psychonomic Bulletin & Review*, *24*(4), 1234–1251. https://doi.org/10.3758/s13423-016-1199-y

Peirce, J. W. (2007). Psychopy—psychophysics software in python. *Journal of Neuroscience Methods*, *162*(1), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pinker, S. (1999). *Words and rules: The ingredients of language*. Weidenfeld & Nicolson.

Pirrone, A., Dickinson, A., Gomez, R., Stafford, T., & Milne, E. (2017). Understanding perceptual judgment in autism spectrum disorder using the drift diffusion model.

Neuropsychology, *31*(2), 173–180. https://doi.org/10.1037/neu0000320

Pober, B. R. (2010). Williams–beuren syndrome. *New England Journal of Medicine*, *362*(3), 239–252. https://doi.org/10.1056/NEJMra0903074

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., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, *111*(1), 159–182. https://doi.org/10.1037/0033-295X.111.1.159

Ratcliff, R., Love, J., Thompson, C. A., & Opfer, J. E. (2012). Children are not like older adults: A diffusion model analysis of developmental changes in speeded responses. *Child Development*, *83*(1), 367–381. https://doi.org/10.1111/j.1467-8624.2011.01683.x

Ratcliff, R., Perea, M., Colangelo, A., & Buchanan, L. (2004). A diffusion model account of normal and impaired readers. *Brain and Cognition*, *55*(2), 374–382. https://doi.org/10.1016/j.bandc.2004.02.051

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*(5), 347–356. https://doi.org/10.1111/1467-9280.00067

Ratcliff, R., Schmiedek, F., & McKoon, G. (2008). A diffusion model explanation of the worst performance rule for reaction time and IQ. *Intelligence*, *36*(1), 10–17. https://doi.org/10.1016/j.intell.2006.12.002

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*(2), 333–367. https://doi.org/10.1037/0033-295X.111.2.333

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging*, *16*(2), 323–341. https://doi.org/10.1037/0882-7974.16.2.323

Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language*, *50*(4), 408–424. https://doi.org/10.1016/J.JML.2003.11.002

Ratcliff, R., Thapar, A., & McKoon, G. (2007). Application of the diffusion model to two-choice tasks for adults 75–90 years old. *Psychology and Aging*, *22*(1), 56–66. https://doi.org/10.1037/0882-7974.22.1.56

Raven, J. C. (1998). *The coloured progressive matrices*. OUP.

Robinson, B. F., Mervis, C. B., & Robinson, B. W. (2003). The roles of verbal short-term memory and working memory in the acquisition of grammar by children with Williams syndrome. *Developmental Neuropsychology*, *23*(1–2), 13–31. https://doi.org/10.1080/87565641.2003.9651885

Scerif, G., Cornish, K., Wilding, J., Driver, J., & Karmiloff-Smith, A. (2004). Visual search in typically developing toddlers and toddlers with fragile X or Williams syndrome. *Developmental Science*, *7*(1), 116–130. https://doi.org/10.1111/j.1467-7687.2004.00327.x

Simos, P. G., Sideridis, G. D., Protopapas, A., & Mouzaki, A. (2011). Psychometric evaluation of a receptive vocabulary test for Greek elementary students. *Assessment for Effective Intervention*, *37*(1), 34–49. https://doi.org/10.1177/1534508411413254

Stone, M. (1960). Models for choice-reaction time. *Psychometrika*, *25*(3), 251–260. https://doi.org/10.1007/BF02289729

Turner, B. M., Forstmann, B. U., Love, B. C., Palmeri, T. J., & van Maanen, L. (2017). Approaches to analysis in model-based cognitive neuroscience. *Journal of Mathematical Psychology*, *76*, 65–79. https://doi.org/10.1016/j.jmp.2016.01.001

Tyler, L. K., Karmiloff-Smith, A., Voice, J. K., Stevens, T., Grant, J., Udwin, O., Davies, M., & Howlin, P. (1997). Do individuals with Williams syndrome have bizarre semantics? Evidence for lexical organization using an on-line task. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *33*(3), 515–527. https://doi.org/10.1016/S0010-9452(08)70233-8

van Maanen, L., Brown, S. D., Eichele, T., Wagenmakers, E.-J., Ho, T. C., Serences, J. T., & Forstmann, B. U. (2011). Neural correlates of trial-to-trial fluctuations in response caution. *Journal of Neuroscience*, *31*(48), 17488–17495. https://doi.org/10.1523/JNEUROSCI.2924-11.2011

van Maanen, L., Forstmann, B. U., Keuken, M. C., Wagenmakers, E.-J., & Heathcote, A. (2016). The impact of MRI scanner environment on perceptual decision-making. *Behavior Research Methods*, *48*(1), 184–200. https://doi.org/10.3758/s13428-015-0563-6

van Maanen, L., & Miletić, S. (2020). The interpretation of behavior-model correlations in unidentified cognitive models. *Psychonomic Bulletin & Review*, https://doi.org/10.3758/s13423-020-01783-y

van Maanen, L., van Rijn, H., & Taatgen, N. (2012). RACE/a: An architectural account of the interactions between learning, task control, and retrieval dynamics. *Cognitive Science*, *36*(1), 62–101. https://doi.org/10.1111/j.1551-6709.2011.01213.x

van Ravenzwaaij, D., Brown, S., & Wagenmakers, E.-J. (2011). An integrated perspective on the relation between response speed and intelligence. *Cognition*, *119*(3), 381–393. https://doi.org/10.1016/J.COGNITION.2011.02.002

Vicari, S., Bates, E., Caselli, M. C., Pasqualetti, P., Gagliardi, C., Tonucci, F., & Volterra, V. (2004). Neuropsychological profile of Italians with Williams syndrome: An example of a dissociation between language and cognition? *Journal of the International Neuropsychological Society*, *10*(6), 862–876. https://doi.org/10.1017/s1355617704106073

Vicari, S., & Carlesimo, G. A. (2006). Short-term memory deficits are not uniform in Down and Williams syndromes. *Neuropsychology Review*, *16*(2), 87–94. https://doi.org/10.1007/s11065-006-9008-4

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology: A practical introduction. *Experimental Psychology*, *60*(6), 385–402. https://doi.org/10.1027/1618-3169/a000218

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, *32*(7), 1206–1220. https://doi.org/10.3758/BF03196893

Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, *114*(3), 830–841. https://doi.org/10.1037/0033-295X.114.3.830

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*(1), 35–57. https://doi.org/10.3758/s13423-017-1343-3

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language*, *58*(1), 140–159. https://doi.org/10.1016/j.jml.2007.04.006

White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010). Using diffusion models to understand clinical disorders. *Journal of Mathematical Psychology*, *54*(1), 39–52. https://doi.org/10.1016/j.jmp.2010.01.004

Winkel, J., van Maanen, L., Ratcliff, R., der Schaaf, M. E., Van Schouwenburg, M. R., Cools, R., & Forstmann, B. U. (2012). Bromocriptine does not alter speed-accuracy tradeoff. *Frontiers in Decision Neuroscience*, *6*. https://doi.org/10.3389/fnins.2012.00126

Withers, S. (1996). A new clinical sign in Williams syndrome. *Archives of Disease in Childhood*, *75*(1), 89. https://doi.org/10.1136/adc.75.1.89

Zeguers, M. H. T., Snellings, P., Tijms, J., Weeda, W. D., Tamboer, P., Bexkens, A., & Huizenga, H. M. (2011). Specifying theories of developmental dyslexia: A diffusion model analysis of word recognition. *Developmental Science*, *14*(6), 1340–1354. https://doi.org/10.1111/j.1467-7687.2011.01091.x