

ARTICLE

Prediction of successful reanalysis based on eye-blink rate and reading times in sentences with local ambiguity

Lola Karsenti^{1*} and Aya Meltzer-Asscher^{1,2} 

¹Linguistics Department, Tel Aviv University, Tel Aviv, Israel; ²Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

*Corresponding author. Email: lolakarsenti@mail.tau.ac.il

(Received 26 January 2022; Revised 08 July 2022; Accepted 18 July 2022)

Abstract

The present study focuses on individual differences in the ability to recover from an initial misinterpretation during the processing of garden path (GP) sentences with local syntactic ambiguity. The performance of reanalysis in GP sentences is a cognitive task that requires efficient use of executive functions and allocation of working memory resources. In this study, we explored the possible role of the neurotransmitter dopamine, which has long been implicated in cognitive control processes, in the successful performance of reanalysis. We examined whether participants' ability to successfully reanalyze a sentence with local ambiguity can be predicted based on (1) their tonic dopamine levels, as reflected by their resting state spontaneous eye-blink rate, measured prior to the experiment; and (2) their reading time patterns in the critical region of the sentence. We ran a self-paced reading experiment in Hebrew, assessing reanalysis performance via a paraphrasing task. We observed a linear and polynomial effect of eye-blink rate on reanalysis performance, with medium rates, corresponding to medium dopamine levels, associated with best performance. We also observed an effect of reading times, with longer reading times in the critical region predicting better reanalysis performance.

Keywords: syntactic processing; sentence comprehension; syntactic reanalysis; dopamine; cognitive control; eye-blink rate; individual differences

1. Introduction

Research in sentence processing aims to develop a model of how humans understand language in real time. Traditionally, this research has been mostly based on the average performance of participant groups. However, such group analyses, even when supplemented with random effects of participants, very likely obscure differences between comprehenders, making it impossible to discover individual effects and processing strategies (e.g., Staub, 2021; Vasisht et al., 2019). Given this, in recent

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

years there is a growing interest in the nature of individual differences in sentence processing (e.g., Cunnings & Fujita, 2021; Freed et al., 2017; James et al., 2018; Johnson & Arnold, 2021; Kim et al., 2018; Novick et al., 2009; Payne & Federmeier, 2019; among others). These studies have shown that language processing is modulated by individual differences in cognitive abilities such as verbal working memory (WM) and cognitive control, as well as by differences in intelligence, language experience, and speed of processing.

In the present study, we focus on individual differences in the ability to recover from an initial misinterpretation during the processing of garden path (GP) sentences, specifically sentences with an object/subject local syntactic ambiguity as in (1):

(1) While Susan drank the water evaporated.

Numerous studies (e.g., Adams et al., 1998; Ferreira & Henderson, 1990; Frazier & Rayner, 1982; Staub, 2007) found longer reading times (RTs) at the disambiguating region, that is, the main verb, in these sentences compared to unambiguous sentences. This pattern was interpreted as reflecting recovery from an initial, ultimately incorrect misparse, in which the ambiguous noun phrase (*the water*) is attached as the object of the first verb. This reanalysis was taken to result in constructing a correct structure for the sentence.

However, these early studies did not look at whether the sentences were ultimately understood correctly. Christianson et al. (2001) first demonstrated that in fact, this is not always the case (see also Christianson et al., 2006; Ferreira et al., 2002; Fujita, 2021; Nakamura & Arai, 2016; Patson et al., 2009; Slattery et al., 2013; van Gompel et al., 2006, among others). For instance, Christianson et al. (2001) found that after reading the sentence *‘While the man hunted the deer ran through the woods’* participants incorrectly answered ‘yes’ to the question *‘Did the man hunt the deer?’* close to 60% of the time. At the same time, participants also correctly answered ‘yes’ to the question *‘Did the deer run through the woods?’* nearly 90% of the time. Based on this, Christianson et al. argued that even when participants were able to partially reanalyze the ambiguous NP as the subject of the main clause, the initial misinterpretation whereby this NP received a thematic role from the embedded verb still lingered. Namely, comprehenders often engage in partial reanalysis and arrive at inaccurate, ‘good enough’ representations based on incomplete processing of the sentence. Recent studies suggest that the lingering of the initial interpretation is attributable not to failure in building a proper structure, but rather to failure in cleaning up all remnants of the syntactic and semantic representations built during the initial misparse (Fujita, 2021; Huang & Ferreira, 2021; Slattery et al., 2013).

Huang and Ferreira (2021) further found that for these locally ambiguous sentences, response accuracy to comprehension questions was negatively correlated with RTs of the disambiguating verb, namely accuracy increased with shorter RTs. This negative correlation was taken by the authors to suggest that the more the parser commits to an initial misparse, the more difficulty it experiences with recovering from that parse, increasing the likelihood of misinterpretation.

In parallel with this line of investigation, a question arises from the perspective of individual differences: Which cognitive abilities are responsible for reanalysis performance? Which participants are (un)able to perform reanalysis? The following subsection elaborates on these issues.

1.1. *Reanalysis, executive function, and individual differences*

It has been argued that the ability to recover from initial misinterpretation of a sentence is supported by cognitive control or executive function, that is, a set of high-level cognitive processes, including response inhibition, conflict monitoring, WM updating, and shifting, that control lower level processes in the service of goal-directed behavior (e.g., Bornkessel et al., 2004; Engelhardt et al., 2017; Hsu et al., 2021; Novick et al., 2014; Vuong & Martin, 2014; Woodard et al., 2016). For example, Novick et al. (2005, 2010) propose that executive function is necessary during recovery from GP in order to resolve the representational conflict created by the contradiction between the initially adopted interpretation and the disambiguating information. Cognitive control plays a crucial role in this process, by suppressing the initially preferred interpretation and by boosting an alternative possible interpretation in accordance with the available information.

Evidence for the role of cognitive control in reanalysis comes from neuropsychological studies showing that lesions in areas supporting cognitive control lead to difficulty with GP sentences (e.g., Novick et al., 2009), as well as from functional magnetic resonance imaging (fMRI) studies (e.g., January et al., 2009) which found engagement of the left inferior frontal gyrus in conflict monitoring both in nonverbal tasks and in comprehension of GP sentences.

Further evidence for cognitive control engagement in syntactic reanalysis comes from behavioral studies looking at individual differences in the neurotypical population. Christianson et al. (2006) have reported negative correlations between reading span scores, reflecting WM capabilities, and GP comprehension errors in older adults. However, when Mendelsohn (2002) tested both WM (using the reading span task) and inhibition (using the Verbal Sorting Task) to compare their predictive value for GP recovery, she found the Verbal Sorting Task to be the best predictor of reanalysis performance abilities. Vuong and Martin (2014) similarly found that verbal inhibitory control plays an important role in performing reanalysis. They showed that RTs of the disambiguating region and successful reanalysis in GP sentences positively correlated with an inhibitory control score from the verbal Stroop task. According to the authors, comprehenders with lower control abilities had greater difficulty in suppressing the initial interpretation. Engelhardt et al. (2017) assessed a large, heterogeneous sample of participants on a battery of cognitive tasks measuring intelligence, speed of processing, inhibitory control and shifting. They found that individuals with higher intelligence and faster processing were more likely to correctly comprehend GP sentences. Inhibition produced only a marginal effect, where individuals with poorer inhibitory control were less likely to answer the comprehension questions correctly.

It can be observed that studies investigating individual differences in language processing often include a battery of cognitive tests measuring WM, executive function, intelligence, speed of processing, and so forth, which are then used as predictors for the linguistic task. However, as noted by many researchers, one problem with this approach is that performance across tasks is almost always correlated (e.g., Long & Freed, 2021). Relatedly, it is often acknowledged that when an effect is attributed to some factor (e.g., WM), there is always a possibility that this factor is a proxy for another, unmeasured predictor (e.g., Nicenboim et al., 2016).

Our study aims to explore a new, neurophysiological measure for predicting individual differences in reanalysis, namely resting-state eye-blink rate (rs-EBR), as

a proxy for dopamine function. Examining dopamine is beneficial, since it is a single predictor which was argued to underlie different cognitive abilities, as explained below. Getting to the primary sources of variation will enable examining less predictors, and will offer more principled explanations for individual differences (see Kurthen et al., 2020, for a similar approach utilizing a different physiological measure, the individual alpha frequency).

1.2. Dopaminergic activity, cognitive function, and eye-blink rate

When investigating the neurobiological correlates of individual differences in language processing and associated cognitive skills, one ‘immediate suspect’ is the activity of the neurotransmitter dopamine (DA) (for reviews on the anatomy and physiology of DA pathways see Cools & D’Esposito, 2011; and Jongkees & Colzato, 2016). Extensive literature shows that WM updating, inhibition and cognitive flexibility are all driven by dopaminergic activity (Cohen et al., 2002; D’Ardenne et al., 2012; Floresco & Magyar, 2006; Hosenbocus & Chahal, 2012; van Schouwenburg et al., 2010; among many others).

There are two mechanisms of dopamine release, phasic and tonic (Goto et al., 2007; Grace, 1991). Phasic dopamine release is stimulus-driven, coding reward as well as prediction error. In addition, WM updating and gating are accompanied by phasic DA bursts (e.g., Hazy et al., 2006; van Schouwenburg et al., 2010). In contrast, tonic dopamine levels represent background dopaminergic activity, which is not stimulus-driven (Frank, 2005; Hernandez-Lopez et al., 1997, as cited in Jongkees & Colzato, 2016). Tonic dopamine levels can be manipulated by drug administration and can be measured by invasive techniques such as positron emission tomography (PET) (e.g., Jongkees & Colzato, 2016). However, an extensive body of research suggests an additional, noninvasive measure of dopaminergic activity. It has been argued that spontaneous, rs-EBR, namely eye-blink rate measured during fixation, in the absence of a cognitive task, is a unique indirect marker of central tonic dopamine function, with higher rs-EBR predicting higher DA function (e.g., Kaminer et al., 2011; Karson, 1983; see review in Jongkees & Colzato, 2016).

The prefrontal basal ganglia WM model (e.g., Frank & O’Reilly, 2006; Hazy et al., 2006) proposes that WM updating and maintenance are coordinated by a dynamic gating mechanism, via two striatal-thalamo-cortical circuits associated with two types of dopamine receptors, D1 and D2. Schematically, D1 dopamine receptors are involved in facilitating WM updating (‘go’ signaling), while D2 dopamine receptors are involved in stable maintenance, suppressing competing responses and representations (see also Cools, 2011). Dopamine has an excitatory effect on D1 signals and an inhibitory effect on D2 signals (for reviews see Goschke & Bolte, 2014; Jongkees & Colzato, 2016; Ott & Nieder, 2019).

Given the characterization above of the relation between DA, updating and maintenance, it is predicted that DA levels will follow an inverted-u-shaped association with performance on tasks requiring cognitive control, rather than following a more-is-better principle (Cools & D’Esposito, 2011; Jongkees, 2020). This is so since, as often noted, cognitive control implies achieving a balance between the opposing demands of stable maintenance of task goals in the face of distractors, and their flexible updating when situational demands have changed (e.g., Miyake et al., 2000; Ueltzhöffer et al., 2015). Importantly, high DA levels can facilitate WM updating up

to a point where it becomes dysfunctional, resulting in heightened distractibility and impaired response inhibition, as the decision threshold is set too low. Conversely, low DA levels may raise the threshold to a point of inducing inflexibility and perseveration (e.g., Dreisbach et al., 2005). Hence, moderate DA levels should be associated with an optimal compromise between stability and flexibility, associated with better performance on many cognitive control tasks (Cools & D'Esposito, 2011).

Indeed, several studies have found that rs-EBR followed an inverted u-shaped association with performance on various cognitive tasks (e.g., Agnoli et al., 2021; Chermahini & Hommel, 2010, 2012; de Rooij & Vromans, 2020). For example, Dang et al. (2016) found an inverted-U-shaped relation between rs-EBR and performance on the anti-saccade task performed after completing the Stroop task. The authors argued that participants with a medium EBR showed reduced costs for switching between the Stroop and anti-saccade task, whereas those with a low or high rs-EBR had less efficient task-switching and thus performed worse on the anti-saccade task.

In addition to task switching, two other key cognitive control processes – inhibitory control and WM updating – were also found to correlate with EBR. For instance, Zhang et al. (2015) showed that increased rs-EBR was correlated with poorer updating of information in WM as found in a visual, letter-based 3-back task. The study also found that increased rs-EBR was related to better shifting and inhibition, as measured by a go/no-go task. This finding contrasts with the findings of Tharp and Pickering (2011) where higher rs-EBR was associated with worse inhibitory control, impairing performance in an incongruent Stroop condition (see also Dreisbach et al., 2005).

Though, as explained above, there is reason to think that DA and cognitive performance should follow an inverted-U-shaped function, the majority of rs-EBR studies (including Zhang et al., 2015 and Tharp & Pickering, 2011 mentioned above) did not look for a polynomial, specifically a quadratic ('parabolic' or U-shaped) relation between the two ($y = x^2$). Instead, most studies only examined the linear correlation between rs-EBR and task performance, or used a median split to distinguish groups of low and high blinkers. As noted by Jongkees and Colzato (2016), such an approach ignores nonlinear patterns in the data, potentially leading to loss of valuable information.

We hypothesize that since rs-EBR reflects the functioning of cognitive control abilities and the updating of WM, it may be a good predictor for individual differences in linguistic tasks demanding cognitive control, specifically, the performance of reanalysis in GP sentences.

1.3. The current study

In the current study, we tested the applicability and usefulness of rs-EBR in psycholinguistic research, and specifically in the examination of individual differences in recovery from GP. We conducted a self-paced reading experiment, asking whether participants' ability to successfully reanalyze a sentence with local ambiguity can be predicted based on (1) their tonic dopamine levels, as reflected by their spontaneous, rs-EBR, measured prior to the experiment; and (2) their RT of the critical region of the sentence.

With regard to the first question above, namely whether reanalysis performance is a function of resting-state EBR (reflecting dopamine levels), we predicted an inverted

U-shaped relation between the two, such that medium blink rates (i.e., medium DA levels) will correspond to better performance, that is, successful reanalysis. This prediction is based on the hypothesis, explained in 1.2 above, that overly low or high dopamine levels are detrimental to tasks requiring cognitive flexibility, WM updating, and inhibition (such as performance of reanalysis), whereas medium dopamine levels are optimal for such tasks.

Our second question was whether RTs in the disambiguating region can predict reanalysis performance. Previous studies of individual differences in performance of reanalysis did not look at RTs, the only exception being Vuong and Martin (2014). However, in the study of Vuong and Martin (2014), only correctly reanalyzed trials were included in the analysis of RTs (and correlated with reaction times in the Stroop task). Therefore, this study does not provide information about the relation between RT of the critical region and accuracy of understanding. Based on previous studies asking similar questions, two opposing predictions can be formulated. One possibility is that participants who read the critical region quickly will show more accurate interpretation, in line with findings from Huang and Ferreira (2021). This may indicate, as proposed by these authors, that longer RTs at the critical regions reflect higher commitment to the initial misparse, resulting in more difficulty in recovery from it. Alternatively, the more accurate participants will read the critical region more slowly. Such a result will be in line with previous studies (investigating other phenomena, i.e., filler-gap dependencies and agreement interference) showing that slower reading is correlated with increased accuracy, perhaps as it enables deeper processing (Laurinavichyute et al., 2017; Nicenboim et al., 2015).

To gauge performance of reanalysis, most previous studies that examined individual differences employed yes/no comprehension questions (Christianson et al., 2001, 2006; Ferreira & Patson, 2007; Huang & Ferreira, 2021; Fujita, 2021; Slattery et al., 2013; but see Vuong & Martin, 2014, who used more open-ended questions). Patson et al. (2009) raised several concerns regarding this methodology. First, it does not force participants to arrive at a final representation of the sentence prior to answering the question. Additionally, the question itself reintroduces the original misinterpretation, thus possibly coercing participants into accepting it. To address this, in Patson et al. (2009), comprehension questions were replaced by a *paraphrasing task*, which forces participants to derive a final interpretation of the sentences they read, without reintroducing the problematic structure. The findings from the paraphrasing task were fully consistent with those of Christianson et al. (2001). In the present study, we measure sentence comprehension (and thereby reanalysis performance) using the paraphrasing task.

2. Methods

2.1. Participants

Ninety-seven native Hebrew speakers, 88 of which were students and the rest young workers at Tel Aviv University's summer camp (mean age = 25.2, range 19–38 years, 59 females) participated in the study in exchange for financial compensation or course credit. All participants gave written informed consent. The study was approved by the local ethics committee (Faculty of Humanities, Tel Aviv University).

2.2. Materials

The experiment included 28 Hebrew sentence sets with seven conditions, two of which we present and analyze in this article (see Appendix A for explanation regarding the additional conditions and why they were not included in the analysis). The relevant conditions for the current article are exemplified in (2). (2a) includes an optionally transitive (OT) verb, *drank*, whereas the baseline sentence (2b) includes an intransitive (IN), unaccusative verb, *woke up*. Based on prior research, we predict that in (2a), the OT verb (*drank*) initially attaches the ambiguous noun phrase (NP) as a direct object, resulting in processing difficulty when the main clause is found to be missing a subject. In contrast, in (2b) the unaccusative verb (*woke up*, marked for intransitivity in Hebrew) has no thematic role to assign to the NP *water*; this NP attaches as the subject of the main verb *flowed*, and no reanalysis is predicted. Sentences were presented region-by-region according to the demarcation shown in (2).

- (2) a. Optionally Transitive verb condition (OT):
axrey she-ha-orxim / šatu / maim / karim / zarmu /
 after that-the-guests / drank / water / cold / flowed /
me-ha-berez / ba-xava.
 from-the-tap / at+the-farm
 ‘After the guests drank cold water flowed from the tap at the farm.’
- b. Intransitive verb condition (IN):
axrey she-ha-orxim / hitoreru / maim / karim / zarmu /
 after that-the-guests / woke-up_{UNACC} / water / cold / flowed /
me-ha-berez / ba-xava.
 from-the-tap / at+the-farm
 ‘After the guests woke up cold water flowed from the tap at the farm.’

For the OT condition, we selected 28 optionally transitive verbs whose proportion of occurrence with a direct object ranges from 20% to 86% (based on manual coding of the first 100 occurrences of each verb in Google). The OT and IN verbs were selected based on a prior self-paced reading experiment, which included the sentences in (2) as well as similar sentences with conjunction (*‘the guests drank / woke up and then water flowed from the tap’*). The results showed no difference in RTs of the verbs between OT and IN verbs (Mean OT = 503 ms (*SD* = 204 ms), Mean IN = 504 ms (*SD* = 225 ms), $p = 0.975$).

In Hebrew, both the subject-verb and the verb-subject word orders are possible, particularly with unaccusative or passive verbs, used in our study as the main verbs. Thus, in (2a), upon the arrival of the main verb, it is still possible that a postverbal NP will serve as its subject (e.g., *axrey she-ha-orxim šatu maim karim zarmu me-ha-berez ba-xava MAIM XAMIM*, ‘After the guests drank cold water flowed from the tap at the farm HOT WATER’). Thus, the realization that the main clause lacks a subject, and the ensuing reanalysis, may happen not at the main verb but rather downstream, at the end of the sentence. Our critical region for RT analysis therefore included the main verb and the sentence-final region.

Sentences were assigned to lists in a Latin Square design. Since the experiment originally included seven conditions, each participant read four locally ambiguous sentences as in (2a) and four baseline sentences as in (2b) (as well as four sentences

each from types (i–v in the Appendix). The number of experimental items per condition was relatively small so as not to encourage careful, strategic reading of the locally ambiguous sentences, while keeping the overall length of the experiment reasonable. The distribution of the 28 experimental sets to lists was done such that each participant read sentences with OT verbs of weak (20–39%), moderate (40–60%) and strong (61–86%) transitivity.

The experimental sentences were intermixed with 72 filler sentences. Thirty filler sentences were similar to conditions (2a) and (2b), but included, in their final interpretation, a transitive verb followed by its direct object (as in: *'After the student submitted all the seminar papers she went on vacation'*), or an intransitive verb following by a PP (as in: *'While the students were concentrating on the long exam the supervisor fell asleep on the chair'*). Other fillers were similar to the other five conditions. Order of presentation was randomized for each participant.

2.3. Procedure

All data were collected between 9:30 am and 4:00 pm, since spontaneous EBR was argued to be stable during daytime (Jongkees & Colzato, 2016). Participants first underwent resting-state EBR measurement, during which they fixated on a cross in the center of a screen for 3.5 minutes, while three electrodes were placed above and below their left eye. The electrooculogram was recorded by a Brain Products experimental system. Individual rs-EBR was calculated during a 3-minute interval (discarding the first 30 seconds of recording) by the *BrainVision Analyzer 2* software.

After this, the SPR portion of the experiment began. Participants first familiarized themselves with the self-paced reading and paraphrasing tasks by completing a practice block with seven trials that did not involve GP sentences. Then, the main experiment started. Each target sentence, as well as some of the filler sentences, were followed by the instruction: 'Write the sentence which you have just read,' without further directions. Participants typed in their responses using the keyboard.

2.4. Paraphrases coding

We identified four categories for paraphrases: successful reanalysis, lingering misinterpretation, inconclusive and incomplete. Paraphrases were coded as reflecting successful reanalysis (R) if they presented one of the following: addition of a comma (see translated example in 3a); switching the order of the clauses (3b); or transferring the ambiguous NP to the post-verbal position of the main clause (3c), a grammatical word order in Hebrew.

- (3) a. After the guests drank, cold water flowed from the tap.
- b. Cold water flowed from the tap after the guests drank.
- c. After the guests drank flowed cold water from the tap.

Paraphrases were coded as reflecting lingering misinterpretation, or partial reanalysis (P), if they indicated that the ambiguous noun phrase was analyzed both as the direct object of the subordinate clause verb and as the subject of the main clause, as in (4a–b).

- (4) a. After the guests drank cold water it flowed from the tap.
 b. After guests drank cold water, cold water flowed from the tap.

Paraphrases were coded as ‘inconclusive’ when they presented word-by-word repetition of the experimental sentence, as in: ‘*After the guests drank cold water flowed from the tap.*’ We believe that such responses may reflect successful reanalysis, partial reanalysis, or perhaps no attempt at reanalysis at all. Finally, when participants could not come up with a paraphrase or when the paraphrase suffered from severe omissions or ungrammaticality, the responses were coded as ‘incomplete.’

2.5. Data analysis

The data of three participants who lacked EBR measurement and of two participants who wrote only paraphrases of the ‘incomplete’ type were excluded from the analysis. Further, since inconclusive responses can reflect either successful or partial reanalysis, our analysis did not include ambiguous responses (an additional analysis which does include ambiguous responses is presented in Appendix B). Twenty-one out of the remaining 92 participants only provided inconclusive responses for the experimental sentences. Therefore, the analysis included data from 71 participants.

We conducted all analyses using R, version 4.0.2 (R Core Team, 2020-06-22). The lme4 package, version 1.1–27.1, was used for regression model analysis (Bates et al. 2015). Throughout our analyses, we used linear and logistic models. In all cases, visual inspection of plots of residuals against fitted values and Q–Q plots revealed no obvious deviations from normality and homoscedasticity. Variance inflation factors were estimated using the car package, version 3.0–10, (Fox & Weisberg, 2019), and no issues with (multi-)collinearity were detected. When fitting interactions, variables were centered and standardized (Winter, 2019). All data and code can be found on the OSF page for the project in https://osf.io/uqvza/?view_only=5f83774ccbe74681b0fa544da3061c2b.

We examined whether the reanalysis performance (RP) of participants is predicted by their resting state eye-blink rate (rs-EBR) and their RT of the critical region (critical RT). The critical region involved three words: the main verb and the sentence final region. To specifically take into account RTs of the critical region, and not the participants’ overall reading rate, the variable *critical RT* represents a subtraction of the individual average RT per word in the filler sentences from the individual average RT per word in the critical region. We used the fillers as a baseline for the participant’s average reading rate, rather than the critical region in the baseline sentences, since different participants saw different baseline sentences, and thus the latter measure would depend on the specific baseline sentences which the participant encountered. Note that different participants also saw different experimental sentences, which could affect the *critical RT* measure. Importantly, however, a one-way ANOVA showed no effect of list on *critical RT* ($p = 0.371$).

As the number of valid, unambiguous (R or P) responses varied between the participants (some had four unambiguous responses, others three, etc.), we used a weighted generalized logistic model (Dobson & Barnett, 2018; chapter 7), that takes into account the number of trials from which the RP score was generated for each participant. The individual RP average score was calculated with $R = 1$ and $P = 0$ across n trials (n between 1 and 4), with n used as the weight in the model. We fitted a

logistic regression model, modeling individual RP score as a function of two continuous variables, namely the participant's rs-EBR (with both linear and polynomial terms, given the literature on an inverse U-shaped relation between EBR and task performance) and the participant's log-transformed average critical RT, as well as their interaction (see Appendix C for the models used in the analyses).

In the model described above, we did not exclude participants based on their blink rate, as the decision of how to define outliers for EBR is not straightforward (see Chermahini & Hommel, 2012). However, the rs-EBR of two participants was found to be above 2.5 standard deviations from the group mean. To test the consequences of the exclusion of these possible outliers, we also fitted the same logistic regression model without the two subjects whose rs-EBR were above 2.5 SD.

To examine the GP effect, we fitted a linear mixed model, modeling the log-transformed average RTs at the critical region, including the main verb and the sentence final region, as a function of condition (GP vs. baseline condition), including random effects for participants and items (see Appendix C for the model).

3. Results

The average RTs at the critical region, including the main verb and the sentence final region, were significantly longer in the GP condition than in the baseline condition (2,687.11 ms, $SD = 1,441.93$ vs. 1,930.49 ms, $SD = 1,025.66$; $p < 0.001$), confirming the robustness of the GP effect in this sentence type in Hebrew, as shown in Fig. 1 (for the details of the model, see Appendix C).

The distribution of the different response types in the two conditions for 92 participants is presented in Table 1. Note that in the baseline condition, R responses do not indicate successful reanalysis (as misinterpretation is not predicted to arise and reanalysis is not required), but rather responses including the same manipulations indicating successful reanalysis in the GP sentences (e.g., addition of a comma, switching of clauses). It may be observed that 'inconclusive' responses, namely word-by-word repetitions of the sentence, made up a considerable part of the responses, in both conditions. This is probably due to the instructions the

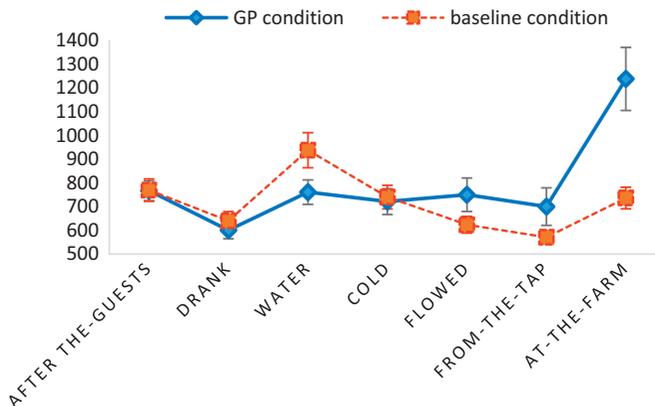


Fig. 1. Mean RTs (all 92 participants) by region in the GP and baseline conditions. Error bars mark ± 1 SE.

Table 1. Paraphrases coding: percentage (number) by category by condition

	Successful (R)	Partial (P)	Inconclusive	Incomplete
GP	26.6% (98)	25.3% (93)	38.6% (142)	9.5% (35)
Baseline	35.6% (131)	0.8% (3)	57.1% (210)	6.5% (24)

participants received, which simply asked them to write the sentence they have just read, without indicating that they should change it in any way.

The relationship between individual rs-EBR and individual average RP for the 71 participants who had no ambiguous paraphrases is shown in Fig. 2 (the figure for all 92 participants can be found in Appendix B). It can be seen in Fig. 2 that participants with rs-EBRs higher than 41 had no successful reanalyses at all.

The relationship between individual rs-EBR, individual critical RT and individual average RP is shown in Table 2. It can be seen that in general, lower RP is associated with higher rs-EBR, and with lower critical RT.

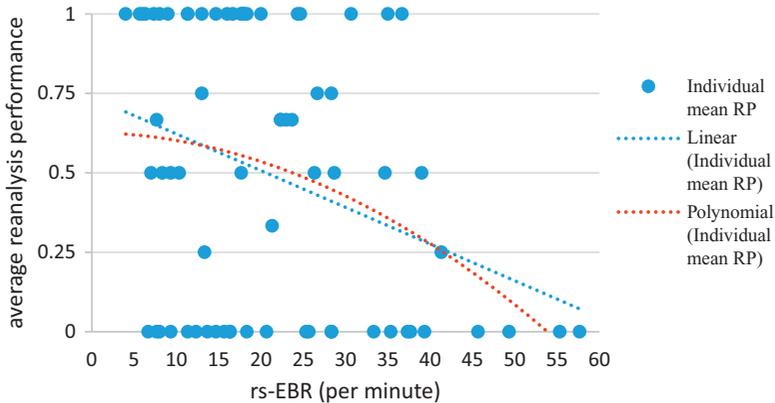


Fig. 2. Individual average reanalysis performance (RP) by rs-EBR, no ambiguous paraphrases.

Table 2. Relationship between rs-EBR, critical RT and individual average RP

Individual average RP	# of participants	rs-EBR (per minute)			Critical RT (ms)	
		Average	Std. Dev.	Range	Average	Std. Dev.
0	11	27.82	15.82	7–55	138.66	190.87
0.125–0.166	8	27.79	16.20	8–58	164.22	149.35
0.25	4	20.58	12.11	11–41	119.69	172.79
0.33–0.375	7	20.19	7.45	12–35	104.35	178.02
0.5	10	19.9	11.02	7–39	109.02	196.13
0.625–0.667	8	16.58	7.41	6–24	75.96	186.80
0.75	8	18.17	7.6	4–28	641.84	470.06
0.83–0.875	7	21.19	11.66	8–37	608.48	942
1	8	14.5	6.3	6–24	644.59	519.31

Note: Data (average RP scores and critical RT) are based on 71 participants and trials with unambiguous (R or P) responses. Critical RT represents a subtraction of the individual average RT per word in the filler sentences from the individual average RT per word in the critical region.

Table 3. Results of regression models

Model	Critical RT	rs-EBR	rs-EBR * critical RT
Model 1: Without exclusion based on rs-EBR (N = 71)	$b = 1.16, t = 4.3,$ $p < .0001$	Linear: $b = -5.38,$ $t = -2.73, p = .006$ Polynomial: $b = -4.76,$ $t = -2.17, p = .029$	$b = -3.5, t = -1.58, p = .11$ $b = 4.4, t = 1.30, p = .19$
Model 2: Possible rs-EBR outliers excluded (N = 69)	$b = .867, t = 4.195,$ $p < .0001$	Linear: $b = -2.286,$ $t = -1.6, p = .1$ Polynomial: $b = -3.664,$ $t = -2.4, p = .015$	$b = -3.733, t = -2.486,$ $p = .013$ $b = 1.479, t = 0.76, p = .44$

The results for the models predicting reanalysis performance based on eye-blink rate and RTs are given in Table 3. For the model without exclusion of outliers, we observed linear and polynomial effects of EBR on RP, with best performance for medium eye blink rates, as well as an effect of critical RT on RP, with longer RTs predicting better reanalysis performance. The interaction between rs-EBR and critical RT was not significant. The predicted values of RP according to the model are visualized in Fig. 3. Similar results were obtained in the model that included inconclusive responses (see Appendix B).

For the model excluding the two participants whose rs-EBR exceeded 2.5 SDs from the group mean, we did not observe a significant linear effect of rs-EBR. In contrast, the polynomial effect was strengthened, and the interaction between the rs-EBR (linear) and RTs became significant, such that for participants with high blink rates, there was less of an effect of RT on reanalysis success.

Finally, in an additional analysis, we examined whether participants' ability to successfully reanalyze is predicted specifically by their slowdown on the critical RT, or whether it is also correlated with their mean reading rate, measured in filler sentences. To perform this analysis, we added to the Model 1 an additional predictor, namely the participant's log-transformed average RT in filler sentences. We found

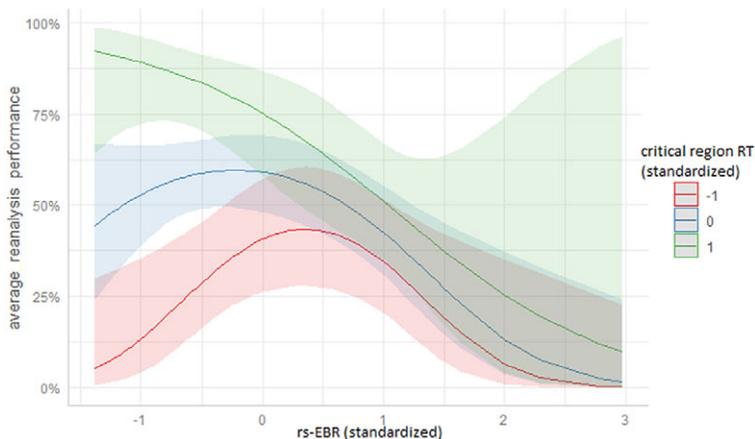


Fig. 3. Predicted values of reanalysis performance by rs-EBR and critical RT for Model 1. Critical RT represents standardized log RT of the critical region (average per word of critical region minus average per word in filler sentences).

that participants' ability to successfully reanalyze was not predicted by their mean reading rate ($p = 0.72$). The other predictors remained significant, with results very similar to those of the original model.

4. Discussion

This study explored individual differences in recovery from initial misanalysis of GP sentences, as indexed by a neurobiological marker for cognitive activity, namely the neurotransmitter dopamine. Cognitive control has a crucial role in recovery from initial misinterpretation of GP sentences (Novick et al., 2005, 2010), and dopamine has long been implicated in cognitive control (Cools & D'Esposito, 2011). As spontaneous eye-blink rate was suggested to be a marker of central dopamine function (Jongkees & Colzato, 2016), in the current study we investigated the relation between resting-state eye-blink rate and recovery from misanalysis. We also examined RTs of the critical region in the sentence, and their relation to reanalysis performance. We assessed reanalysis performance via a paraphrasing task.

The interpretations that participants provided for the GP sentences in the current study were similar to those that Patson et al. (2009) reported for English speakers in the paraphrasing task. Just as in that study, our Hebrew-speaking participants used the two patterns we were interested in: successful reanalysis and lingering of the misinterpretation of the ambiguous NP as an object. It can be noted that in our experiment, these two paraphrase types appeared in similar rates (~25%), whereas Patson et al. report a much higher rate of partial reanalyses (69%). Notably, however, Patson et al. also reported a very high rate (38%) of partial reanalysis paraphrases for their unambiguous baseline sentences (with a comma following the first verb). We believe that there is a difference in the materials of the two studies, which can at least partly explain the difference in the results. As noted by Nakamura and Arai (2016), the high rate of partial reanalyses even in unambiguous sentences in the Patson et al. experiment indicates that comprehenders made inferences about the implicit argument of the verb (i.e., *the boat in While the skipper sailed(,) the boat veered off course*) based on pragmatic information, even without structural ambiguity. Indeed, it can be noted that in the Patson et al. materials, the main clause subject can be very plausibly simultaneously interpreted as the object for the embedded optionally transitive verb in all sentences; in fact, this is the most plausible interpretation of the sentences. In contrast, in about half of the 28 sets in our experiment, such an inference is impossible or implausible, for example, in *'After the team played two extra games were canceled by the club authorities,'* it is not very probable that two extra games were canceled after two extra games were played.

Importantly, we found that successful performance of reanalysis depends on the participant's tonic DA level. In the model with no outlier exclusion, we observed a linear relation, such that higher rs-EBR was associated with lower reanalysis performance. This relation was supplemented, in both models, by a polynomial relation between reanalysis performance and EBR, with best performance for medium DA levels. In addition, we found that longer RTs of the critical region predict successful reanalysis. In the following subsections, we discuss these findings in more detail.

4.1. Reanalysis performance and tonic dopamine levels

The relationship we found between DA and reanalysis performance is consistent with the model of tonic DA function offered by Cools and D'Esposito (2011) for tasks requiring cognitive control. As explained in the Introduction, DA has an excitatory effect on D1 signals, which facilitate WM updating, and an inhibitory effect on D2 signals, which support suppression of competing responses. The combined effect of high DA levels is thus heightened distractibility, to the point of impaired inhibition, whereas the effect of low DA levels is inflexibility and perseveration (Jongkees & Colzato, 2016). In our results, we found a linear relation between EBR and reanalysis performance, such that participants with high DA levels performed worse on reanalysis, confirming Cools & D'Esposito's 2011 observation that a 'more is better' principle is not the correct characterization for the association of DA with performance in cognitive control tasks. In particular, one notable finding in our study is that the group of participants with the highest rs-EBR could not perform successful reanalysis in any trial and therefore did not understand any of the GP sentences correctly. This is to be expected, if these participants exhibit impaired inhibition, and are thus not suppressing the initial interpretation of the sentence, giving rise to partial reanalysis. The behavior of this group of participants is in line with the characterization of partial reanalysis as failure in cleaning up remnants of the syntactic and semantic representations built during the initial misparse (e.g., Fujita, 2021; Huang & Ferreira, 2021; Slattery et al., 2013).

According to the model by Cools and D'Esposito (2011), the lowest levels of DA should not be optimal for reanalysis performance either, as participants with low DA exhibit less cognitive flexibility and heightened perseveration, making it harder to adopt a new analysis. Participants with moderate DA levels are predicted to perform best on the task, since with these DA levels, updating to a new analysis is possible, but suppression of the previous interpretation can also take place. Indeed, in line with this prediction, our models showed a U-shaped relation between eye-blink rate and performance of reanalysis.

Interestingly, in our findings, the results for individuals with low tonic dopamine levels were not clear. Some low DA participants indeed had difficulty with reanalysis, which may reflect difficulty with flexibly finding and updating a new and correct analysis. However, other low DA participants performed successfully, making the polynomial relation between EBR and reanalysis performance weak. The heterogeneous nature of the low DA group can be explained by appealing to other traits and processes indexed by dopamine, namely motivation and learning (Bromberg-Martin et al., 2010; Mohebi et al., 2019; Wise, 2004). Cools (2019) proposes that tonic DA predicts motivation, whereas phasic DA is related to reward learning. We believe that both motivation and reward learning are relevant in the current experiment. A high level of motivation was required in order to perform well in the present experiment, which was long and difficult. On the other hand, feedback provided by participants during debriefing after the experiment showed that the experiment was rewarding and interesting for 'language lovers' and participants who like to excel in difficult tasks. We tentatively suggest the possibility that the heterogeneity of the low rs-EBR participants' performance can be viewed as reflecting differences in participants' reinforcement learning during the experiment: participants whose motivation was initially low, as reflected by low tonic DA before the experiment, could elevate their levels of DA through reward learning during the experiment, or not. Participants

whose DA was elevated to the optimal level subsequently succeeded in reanalysis performance.

It is interesting to consider this hypothesis regarding low DA participants in light of previous studies on the relation between rs-EBR and cognitive flexibility. Chermahini and Hommel (2010) found an inverted U-shaped function relating preexperimental EBR to cognitive flexibility and creativity, with best performance for individuals with medium EBR. In a later study, Chermahini and Hommel (2012) used a mood-induction task to examine the hypothesis that positive mood may improve creativity and flexibility, and that this relation is mediated by dopamine. EBR was measured before and after the task. The authors found that in individuals with preexperimentally low EBR the positive mood-inducing task improved cognitive flexibility and elevated the posttask EBR. In individuals with preexperimentally high EBR, the mood-inducing task did not change cognitive flexibility. It is therefore possible that for low EBR participants, the performance during an experiment is more susceptible to modulation by motivation or rewards, while this is not the case for high EBR participants. In future research, the heterogeneous behavior of low EBR participants can be further investigated, possibly with an additional predictor, such as postexperimental or phasic EBR. The heterogeneity of the low rs-EBR participants tallies with the characterization of rs-EBR in Chermahini and Hommel (2010) and Agnoli et al. (2021) as a very basic measure of DA, which does not distinguish between different dopaminergic pathways and receptors systems.

4.2. Reanalysis performance and reading times of the critical region

Our mean RT analysis revealed a clear processing difficulty in the GP sentences compared to the baseline sentences, which showed up most clearly in the sentence-final region. Regarding individual differences, our findings indicate that participants who slowed down in the face of processing difficulty were able to perform successful reanalysis. For instance, RTs of the critical region were about 2 SDs higher than group's average in the participants that performed successful reanalysis 75–100% of the time. Moreover, in the supplementary analysis we found that participants' ability to successfully reanalyze was not correlated with their mean reading rate, measured as reading speed in filler sentences. Thus, participants who specifically did not slow down in the critical region, did not perform successful reanalysis. These results comply with the well-known generalization that increase in processing speed usually comes with accuracy deterioration (namely the speed-accuracy trade-off).

Our findings are in line with Vuong and Martin (2014), who found that less revision time was associated with poorer cognitive control (measured in a verbal Stroop task). Interpreting this pattern, Vuong and Martin (2014) argue that poorer control readers have 'a tendency to go with good-enough interpretations,' skipping reanalysis when the demand for cognitive control increases. Unfortunately, this study does not offer information with regard to the relation between RT of the critical region and accuracy of understanding. Similarly, MacDonald et al. (1992) and Pearlmutter and MacDonald (1995) investigated the role of individual differences in the processing of sentences with main verb/reduced relative ambiguity, and found that whereas low memory span readers showed no differences in RTs between ambiguous and unambiguous sentences, high memory span readers had reliably

longer RTs at the points of disambiguation in the ambiguous sentences. However, in these studies too, comprehension was not assessed.

Huang and Ferreira (2021) did consider comprehension accuracy against RTs of the critical region, though not looking at individual differences. In this study, response accuracy to the comprehension question in the self-paced reading experiment was negatively correlated with the RTs of the disambiguating verb, such that trials with inaccurate responses were associated with longer RTs than trials with accurate responses. These results contrast with our current results. Note, that in Huang and Ferreira (2021), the RTs difference between the ambiguous and unambiguous conditions at the disambiguating regions was in the order of 10 ms, whereas other experiments show that reanalysis takes longer. Coupled with the low accuracy rate on unambiguous conditions (64%) and the inclusion of a high number (240) of long and difficult sentences, it is possible that Huang & Ferreira's experiment was challenging for participants, leading them to opt for skipping reanalysis in the critical region.

In contrast to Huang and Ferreira (2021), the current study and several others found that correct comprehension corresponded to longer RTs. For instance, Blott et al. (2021) found that successful recovery from misinterpretation caused by another type of ambiguity, namely semantic ambiguity, was associated with a large processing cost (400 ms), such that the comprehenders who judged the experimental sentences correctly took longer to read the ambiguous as opposed to unambiguous sentences. Laurinavichyute et al. (2017), in a study of reflexive processing in Russian, similarly found that more accurate participants (as verified in yes/no comprehension questions) read the retrieval site more slowly in sentences with similarity-based inference compared to those without interference.

Interestingly, Nicenboim et al. (2016) and Van Dyke et al. (2014) demonstrate that when we look carefully at individual abilities of participants, the results become more informative than a simple speed-accuracy trade-off. Specifically, Van Dyke et al. (2014) examined individual differences in susceptibility to retrieval interference and found that higher accuracy was coupled with slow reading and characterized high-span participants, whereas miscomprehension was coupled with faster reading and characterized low-span participants. Nicenboim et al. (2016) reported that WM capacity was found to correlate with comprehension accuracy, and slowdowns were attested only in high-capacity readers, leading the authors to suggest that 'in some cases, interpreting longer RTs as indexing increased processing difficulty and shorter RTs as facilitation may be too simplistic...' (p. 21). Nicenboim et al. (2016) and Van Dyke et al. (2014) further argue that the predictor for comprehension is not a WM capacity per se, but rather, WM capacity is a proxy to other resources or abilities. Nevertheless, the suggestion behind these findings is that better comprehenders slow down during difficulty because they have the abilities or resources with which to handle the difficulty successfully (e.g., Kim et al., 2018; Nicenboim et al., 2016).

Interestingly, we also observed a significant interaction between RTs and rs-EBR in predicting reanalysis performance in the model excluding the two outliers, and a similar pattern in the model with all participants, as can also be observed in Fig. 3, showing this model's predictions. The interaction shows that longer RTs improve reanalysis performance only in participants with low and moderate rs-EBR. For high rs-EBR participants, their reanalysis performance is predicted to be very low, regardless of RTs (in practice, slowdowns were not attested in high rs-EBR participants). We suggest tentatively (due to the small number of participants with very high

rs-EBR), that inability to slow down is a source of considerable difficulty for participants with high rs-EBR. It can be noted that the interaction between rs-EBR and RTs may have not turned out significant in the first model due to the study not being sufficiently powered, as a consequence of the small number of items per participant per condition.

5. Conclusion

In the current experiment, we contributed to the research of individual differences in reanalysis performance, presenting an exploratory methodological study on the role of dopamine function in sentence processing. We found that successful reanalysis performance depends on RTs of the critical region, and on resting state EBR, a proxy of striatal dopamine.

The relations between cognitive function, dopamine, and eye-blink rate are a topic for ongoing research and debates (Broadway et al., 2018; Cools, 2019; Tan & Hagoort, 2020), and it is important to keep in mind that EBR provides only a very basic measure of DA. Still, EBR is an accessible, noninvasive measure, providing a useful proxy of DA function, and we propose that examining the role of dopamine in language comprehension may offer insight into individual differences, particularly since, as emphasized in James et al. (2018) and Kim et al. (2018), there is likely no single cognitive ability that predicts sentence processing outcomes.

Acknowledgments. We thank Tal Ness for her help with using Brain Products equipment and software. We thank audiences at the CUNY Conference on Human Sentence Processing 33 and 34, for their valuable feedback. Lastly, we thank the editor and two anonymous reviewers for insightful comments on earlier versions of this article, which has helped us improve the paper substantially.

Data availability statement. All data and R code are available in the following OSF online repository: https://osf.io/uoyvza/?view_only=5f83774cbe74681b0fa544da3061c2b.

Competing interests. The authors declare none.

Funding statement. The research presented in this article was funded by the Israel Science Foundation, grant number 216/18 (P.I. A.M.-A.).

References

- Adams, B. C., Clifton Jr., C., & Mitchell, D. C. (1998). Lexical guidance in sentence processing? *Psychonomic Bulletin & Review*, 5(2), 265–270. <https://doi.org/10.3758/BF03212949>
- Agnoli, S., Mastria, S., Zanon, M., & Corazza, G. E. (2021). Dopamine supports idea originality: The role of spontaneous eye blink rate on divergent thinking. *Preprint*. <https://doi.org/10.31234/osf.io/rfgu4>
- Bates D., Mächler M., Bolker B., Walker S. (2015). “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software*, 67(1), 1–48. doi:10.18637/jss.v067.i01.
- Blott, L. M., Rodd, J. M., Ferreira, F., & Warren, J. E. (2021). Recovery from misinterpretations during online sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(6), 968–997. <https://doi.org/10.1037/xlm0000936>
- Bornkessel, I. D., Fiebach, C. J., & Friederici, A. D. (2004). On the cost of syntactic ambiguity in human language comprehension: An individual differences approach. *Cognitive Brain Research*, 21(1), 11–21. <https://doi.org/10.1016/j.cogbrainres.2004.05.007>
- Broadway, J. M., Frank, M. J., & Cavanagh, J. F. (2018). Dopamine D2 agonist affects visuospatial working memory distractor interference depending on individual differences in baseline working memory span. *Cognitive, Affective, & Behavioral Neuroscience*, 18, 509–520. <https://doi.org/10.3758/s13415-018-0584-6>

- Bromberg-Martin, E. S., Matsumoto, M., & Hikosaka, O. (2010). Dopamine in motivational control: Rewarding, aversive, and alerting. *Neuron*, 68(5), 815–834. <https://doi.org/10.1016/j.neuron.2010.11.022>
- Chermahini, S. A., & Hommel, B. (2010). The (b)link between creativity and dopamine: Spontaneous eye blink rates predict and dissociate divergent and convergent thinking. *Cognition*, 115(3), 458–465. <https://doi.org/10.1016/j.cognition.2010.03.007>
- Chermahini, S. A., & Hommel, B. (2012). More creative through positive mood? Not everyone! *Frontiers in Human Neuroscience*, 6(12), 319. <https://doi.org/10.3389/fnhum.2012.00319>
- Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, 42(4), 368–407. <https://doi.org/10.1006/cogp.2001.0752>
- Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006). Younger and older adults' "good-enough" interpretations of garden-path sentences. *Discourse Processes*, 42(2), 205–238. https://doi.org/10.1207/s15326950dp4202_6
- Cohen, J. D., Braver, T. S., & Brown, J. W. (2002). Computational perspectives on dopamine function in prefrontal cortex. *Current Opinion in Neurobiology*, 12(2), 223–229. [https://doi.org/10.1016/S0959-4388\(02\)00314-8](https://doi.org/10.1016/S0959-4388(02)00314-8)
- Cools, R. (2011). Dopaminergic control of the striatum for high-level cognition. *Current Opinion in Neurobiology*, 21(3), 402–407. <https://doi.org/10.1016/j.conb.2011.04.002>
- Cools, R. (2019). Chemistry of the adaptive mind: Lessons from dopamine. *Neuron*, 104(1), 113–131. <https://doi.org/10.1016/j.neuron.2019.09.035>
- Cools, R., & D'Esposito, M. (2011). Inverted-U-shaped dopamine actions on human working memory and cognitive control. *Biological Psychiatry*, 69(12), e113–e125. <https://doi.org/10.1016/j.biopsych.2011.03.028>
- Cummings, I., & Fujita, H. (2021). Quantifying individual differences in native and nonnative sentence processing. *Applied Psycholinguistics*, 42(3), 579–599. <https://doi.org/10.1017/S0142716420000648>
- D'Ardenne, K., Eshel, N., Luka, J., Lenartowicz, A., Nystrom, L. E., & Cohen, J. D. (2012). Role of prefrontal cortex and the midbrain dopamine system in working memory updating. *Proceedings of the National Academy of Sciences*, 109(49), 19900–19909. <https://doi.org/10.1073/pnas.1116727109>
- Dang, J., Xiao, S., Liu, Y., Jiang, Y., & Mao, L. (2016). Individual differences in dopamine level modulate the ego depletion effect. *International Journal of Psychophysiology*, 99, 121–124. <https://doi.org/10.1016/j.ijpsycho.2015.11.013>
- de Rooij, A., & Vromans, R. (2020). The (dis) pleasures of creativity: Spontaneous eye blink rate during divergent and convergent thinking depends on individual differences in positive and negative affect. *The Journal of Creative Behavior*, 54(2), 436–452. <https://doi.org/10.1002/jocb.379>
- Dobson, A. J., & Barnett, A. G. (2018). *An introduction to generalized linear models*, 4th ed. Chapman and Hall/CRC. <https://doi.org/10.1201/9781315182780>
- Dreisbach, G., Müller, J., Goschke, T., Strobel, A., Schulze, K., Lesch, K. P., & Brocke, B. (2005). Dopamine and cognitive control: The influence of spontaneous eyeblink rate and dopamine gene polymorphisms on perseveration and distractibility. *Behavioral Neuroscience*, 119(2), 483. <https://doi.org/10.1037/0735-7044.119.2.483>
- Engelhardt, P. E., Nigg, J. T., & Ferreira, F. (2017). Executive function and intelligence in the resolution of temporary syntactic ambiguity: An individual differences investigation. *Quarterly Journal of Experimental Psychology*, 70(7), 1263–1281. <https://doi.org/10.1080/17470218.2016.1178785>
- Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11(1), 11–15. <https://doi.org/10.1111/1467-8721.00158>
- Ferreira, F., & Henderson, J. M. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(4), 555–568. <https://doi.org/10.1037/0278-7393.16.4.555>
- Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, 1(1–2), 71–83. <https://doi.org/10.1111/j.1749-818X.2007.00007.x>
- Floresco S. B., & Magyar, O. (2006). Mesocortical dopamine modulation of executive functions: beyond working memory. *Psychopharmacology (Berl)*, 188(4), 567–585. <https://doi.org/10.1007/s00213-006-0404-5>

- Fox J., Weisberg S. (2019). *An R Companion to Applied Regression*, Third edition. Sage, Thousand Oaks CA. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>.
- Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and non-medicated Parkinsonism. *Journal of Cognitive Neuroscience*, 17, 51–72.
- Frank, M. J., & O'Reilly, R. C. (2006). A mechanistic account of striatal dopamine function in human cognition: Psychopharmacological studies with cabergoline and haloperidol. *Behavioral Neuroscience*, 120(3), 497–517. <https://doi.org/10.1037/0735-7044.120.3.497>
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14, 178–210. [https://doi.org/10.1016/0010-0285\(82\)90008-1](https://doi.org/10.1016/0010-0285(82)90008-1)
- Freed, E. M., Hamilton, S. T., & Long, D. L. (2017). Comprehension in proficient readers: The nature of individual variation. *Journal of Memory and Language*, 97, 135–153. <https://doi.org/10.1016/j.jml.2017.07.008>
- Fujita, H. (2021). On the parsing of garden-path sentences. *Language, Cognition and Neuroscience*, 36(10), 1234–1245. <https://doi.org/10.1080/23273798.2021.1922727>
- Goschke, T., & Bolte, A. (2014). Emotional modulation of control dilemmas: The role of positive affect, reward, and dopamine in cognitive stability and flexibility. *Neuropsychologia*, 62, 403–423. <https://doi.org/10.1016/j.neuropsychologia.2014.07.015>
- Goto, Y., Otani, S., & Grace, A. A. (2007). The Yin and Yang of dopamine release: A new perspective. *Neuropharmacology*, 53(5), 583–587. <https://doi.org/10.1016/j.neuropharm.2007.07.007>
- Grace, A. A. (1991). Phasic versus tonic dopamine release and the modulation of dopamine system responsivity: A hypothesis for the etiology of schizophrenia. *Neuroscience*, 41(1), 1–24. [https://doi.org/10.1016/0306-4522\(91\)90196-u](https://doi.org/10.1016/0306-4522(91)90196-u)
- Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2006). Banishing the homunculus: making working memory work. *Neuroscience*, 139(1), 105–118. <https://doi.org/10.1016/j.neuroscience.2005.04.067>
- Hernández-López, S., Bargas, J., Surmeier, D. J., Reyes, A., Galarraga, E., (1997). D1 receptor activation enhances evoked discharge in neostriatal medium spiny neurons by modulating an L-type Ca²⁺ conductance. *Journal of Neuroscience*, 17, 3334–3342.
- Hosenbocus, S., & Chahal, R. (2012). A review of executive function deficits and pharmacological management in children and adolescents. *Journal of the Canadian Academy of Child and Adolescent Psychiatry = Journal de l'Académie canadienne de psychiatrie de l'enfant et de l'adolescent*, 21(3), 223–229. Retrieved from <https://pubmed.ncbi.nlm.nih.gov/22876270>.
- Hsu, N. S., Kuchinsky, S. E., & Novick, J. M. (2021). Direct impact of cognitive control on sentence processing and comprehension. *Language, Cognition and Neuroscience*, 36(2), 211–239. <https://doi.org/10.1080/23273798.2020.1836379>
- Huang, Y., & Ferreira, F. (2021). What causes lingering misinterpretations of garden-path sentences: Incorrect syntactic representations or fallible memory processes? *Journal of Memory and Language*, 121, 104288. <https://doi.org/10.1016/j.jml.2021.104288>
- James, A. N., Fraundorf, S. H., Lee, E. K., & Watson, D. G. (2018). Individual differences in syntactic processing: Is there evidence for reader-text interactions? *Journal of Memory and Language*, 102, 155–181. <https://doi.org/10.1016/j.jml.2018.05.006>
- January, D., Trueswell, J. C., & Thompson-Schill, S. L. (2009). Co-localization of stroop and syntactic ambiguity resolution in Broca's area: Implications for the neural basis of sentence processing. *Journal of Cognitive Neuroscience*, 21(12), 2434–2444. <https://doi.org/10.1162/jocn.2008.21179>
- Johnson, E., & Arnold, J. E. (2021). Individual differences in print exposure predict use of implicit causality in pronoun comprehension and referential prediction. *Frontiers in Psychology*, 12, 672109. <https://doi.org/10.3389/fpsyg.2021.672109>
- Jongkees, B. J. (2020). Baseline-dependent effect of dopamine's precursor L-tyrosine on working memory gating but not updating. *Cognitive, Affective & Behavioral Neuroscience*, 20(3), 521–535. <https://doi.org/10.3758/s13415-020-00783-8>
- Jongkees, B. J., & Colzato, L. S. (2016). Spontaneous eye blink rate as predictor of dopamine-related cognitive function—A review. *Neuroscience and Biobehavioral Reviews*, 71, 58–82. <https://doi.org/10.1016/j.neubiorev.2016.08.020>

- Kaminer, J., Powers, A. S., Horn, K. G., Hui, C., & Evinger, C. (2011). Characterizing the spontaneous blink generator: An animal model. *Journal of Neuroscience*, 31(31), 11256–11267. <https://doi.org/10.1523/JNEUROSCI.6218-10.2011>
- Karson, C. N. (1983). Spontaneous eye-blink rates and dopaminergic systems. *Brain*, 106(3), 643–653. <https://doi.org/10.1093/brain/106.3.643>
- Kim, A. E., Oines, L., & Miyake, A. (2018). Individual differences in verbal working memory underlie a tradeoff between semantic and structural processing difficulty during language comprehension: An ERP investigation. *Journal of Experimental Psychology: Learning Memory and Cognition*, 44(3), 406–420. <https://doi.org/10.1037/xlm0000457>
- Kurthen, I., Meyer, M., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2020). Individual differences in peripheral hearing and cognition reveal sentence processing differences in healthy older adults. *Frontiers in Neuroscience*, 14, 573513. <https://doi.org/10.3389/fnins.2020.573513>
- Laurinavichyute, A., Jäger, L. A., Akinina, Y., Roß, J., & Dragoy, O. (2017). Retrieval and encoding interference: Cross-linguistic evidence from anaphor processing. *Frontiers in Psychology*, 8, 1–18. <https://doi.org/10.3389/fpsyg.2017.00965>
- Long, D. L., & Freed, E. M. (2021). An individual differences examination of the relation between reading processes and comprehension. *Scientific Studies of Reading*, 25(2), 104–122. <https://doi.org/10.1080/10888438.2020.1748633>
- MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, 24, 56–98.
- Mendelsohn, A. (2002). *Individual differences in ambiguity resolution: Working memory and inhibition*. Unpublished doctoral dissertation, Northeastern University. Retrieved from <https://web.northeastern.edu/npearlmutter/students/papers/aamth.pdf>.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49–100. <https://doi.org/10.1006/cogp.1999.0734>
- Mohebi, A., Pettibone, J. R., Hamid, A. A., Wong, J. T., Vinson, L. T., Patriarchi, T., Tian, L., Kennedy, R. T., & Berke, J. D. (2019). Dissociable dopamine dynamics for learning and motivation. *Nature*, 570, 65–70.
- Nakamura, C., & Arai, M. (2016). Persistence of initial misanalysis with no referential ambiguity. *Cognitive Science*, 40(4), 909–940. <https://doi.org/10.1111/cogs.12266>
- Nicenboim, B., Engelmann, F., Suckow, K., and Vasishth, S. (2015). Fail fast or succeed slowly: good-enough processing can mask interference effects. In N. A. Taatgen, M. K. van Vugt, J. P. Borst, & K. Mehlhorn (Eds.), *Proceedings of ICCM 2015: 13th International Conference on Cognitive Modeling: April 9–11, Groningen, The Netherlands* (pp. 196–197). Groningen: University of Groningen.
- Nicenboim, B., Logačev, P., Gattei, C., & Vasishth, S. (2016). When high-capacity readers slow down and low-capacity readers speed up: Working memory and locality effects. *Frontiers in Psychology*, 7, 280. <https://doi.org/10.3389/fpsyg.2016.00280>
- Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., & Bunting, M. F. (2014). Clearing the garden-path: Improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience*, 29(2), 186–217. <https://doi.org/10.1080/01690965.2012.758297>
- Novick, J. M., Kan, I. P., Trueswell, J. C., & Thompson-Schill, S. L. (2009). A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive Neuropsychology*, 26(6), 527–567. <https://doi.org/10.1080/02643290903519367>
- Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca’s area in sentence comprehension. *Cognitive, Affective and Behavioral Neuroscience*, 5(3), 263–281. <https://doi.org/10.3758/CABN.5.3.263>
- Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2010). Broca’s area and language processing: Evidence for the cognitive control connection. *Language and Linguistics Compass*, 4(10), 906–924. <https://doi.org/10.1111/j.1749-818X.2010.00244.x>
- Ott, T., & Nieder, A. (2019). Dopamine and cognitive control in prefrontal cortex. *Trends in Cognitive Sciences*, 23(3), 213–234. <https://doi.org/10.1016/j.tics.2018.12.006>
- Patson, N. D., Darowski, E. S., Moon, N., & Ferreira, F. (2009). Lingering misinterpretations in garden-path sentences: Evidence from a paraphrasing task. *Journal of Experimental Psychology: Learning Memory and Cognition*, 35(1), 280–285. <https://doi.org/10.1037/a0014276>

- Payne, B., & Federmeier, K. D. (2019). Individual differences in reading speed are linked to variability in the processing of lexical and contextual information: Evidence from single-trial event-related brain potentials. *Word*, 65(4), 252–272. <https://doi.org/10.1080/00437956.2019.1678826>
- Pearlmuter, N. J., & MacDonald, M. C. (1995). Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory and Language*, 34(4), 521–542. <https://doi.org/10.1006/jmla.1995.1024>
- Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, 69(2), 104–120. <https://doi.org/10.1016/j.jml.2013.04.001>
- Staub, A. (2007). The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning Memory and Cognition*, 33(3), 550–569. <https://doi.org/10.1037/0278-7393.33.3.550>
- Staub, A. (2021). How reliable are individual differences in eye movements in reading? *Journal of Memory and Language*, 116, 104190. <https://doi.org/10.1016/j.jml.2020.104190>
- Tan, Y., & Hagoort, P. (2020). Catecholaminergic modulation of semantic processing in sentence comprehension. *Cerebral Cortex*, 30(12), 6426–6443. <https://doi.org/10.1093/cercor/bhaa204>
- Tharp, I. J., & Pickering, A. D. (2011). Individual differences in cognitive-flexibility: The influence of spontaneous eyeblink rate, trait psychoticism and working memory on attentional set-shifting. *Brain and Cognition*, 75(2), 119–125. <https://doi.org/10.1016/j.bandc.2010.10.010>
- Ueltzhöffer, K., Armbruster-Genç, D. J., & Fiebach, C. J. (2015). Stochastic dynamics underlying cognitive stability and flexibility. *PLoS Computational Biology*, 11(6), e1004331. <https://doi.org/10.1371/journal.pcbi.1004331>
- Van Dyke, J. A., Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition*, 131(3), 373–403. <https://doi.org/10.1016/j.cognition.2014.01.007>
- van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Jacob, G. (2006). The activation of inappropriate analyses in garden-path sentences: Evidence from structural priming. *Journal of Memory and Language*, 55(3), 335–362. <https://doi.org/10.1016/j.jml.2006.06.004>
- van Schouwenburg, M., Aarts, E., & Cools, R. (2010). Dopaminergic modulation of cognitive control: Distinct roles for the prefrontal cortex and the basal ganglia. *Current Pharmaceutical Design*, 16(18), 2026–2032. <https://doi.org/10.2174/138161210791293097>
- Vasishth, S., Nicenboim, B., Engelmann, F., & Burchert, F. (2019). Computational models of retrieval processes in sentence processing. *Trends in Cognitive Sciences*, 23(11), 968–982. <https://doi.org/10.1016/j.tics.2019.09.003>
- Vuong, L. C., & Martin, R. C. (2014). Domain-specific executive control and the revision of misinterpretations in sentence comprehension. *Language, Cognition and Neuroscience*, 29(3), 312–325. <https://doi.org/10.1080/01690965.2013.836231>
- Winter, B. (2019). *Statistics for Linguists: An introduction using R*, 1st ed. Routledge. <https://doi.org/10.4324/9781315165547>
- Wise R. A. (2004). Dopamine, learning and motivation. *Nature Reviews: Neuroscience*, 5(6), 483–494. <https://doi.org/10.1038/nrn1406>
- Woodard, K., Pozzan, L., & Trueswell, J. C. (2016). Taking your own path: Individual differences in executive function and language processing skills in child learners. *Journal of Experimental Child Psychology*, 141, 187–209. <https://doi.org/10.1016/j.jecp.2015.08.005>
- Zhang, T., Mou, D., Wang, C., Tan, F., Jiang, Y., Lijun, Z., & Li, H. (2015). Dopamine and executive function: Increased spontaneous eye blink rates correlate with better set-shifting and inhibition, but poorer updating. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, 96(3), 155–161. <https://doi.org/10.1016/j.ijpsycho.2015.04.010>

A. Appendix A: Additional conditions in the experiment

The other five conditions which were originally included in the experiment are exemplified (translated to English) in (i–v).

- (i) The owner brought to the guests that drank cold water last night at the farm.
- (ii) The owner brought to the guests that woke up cold water last night at the farm.

- (iii) The owner brought to the guests that drank cold water orange juice last night at the farm.
- (iv) The owner brought to the guests that got cold water orange juice last night at the farm.
- (v) The owner brought to the guests that showered with cold water orange juice last night at the farm.

We originally intended to test performance of reanalysis in these sentence types as well. However, coding the paraphrases for these five conditions turned out to be problematic, since for the majority of cases we could not determine whether participants successfully reanalyzed the sentence or not. For example, in condition (i), most responses were word-by-word repetitions of the sentence (see Section 2.4). In Hebrew, this sentence type does not allow a comma after the relative clause. Thus, the only response type unambiguously showing correct interpretation would include fronting the ambiguous NP, as in: 'The owner brought cold water to the guests that drank.' However, this was a very rare response pattern. Additionally, for these conditions paraphrases exhibited many omissions and replacements, possibly a result of the memory load arising from the sentences' length, and it was difficult to decide how to code such responses. For example, many responses to condition (i) lacked the subordinate verb, as in: 'The owner brought to the guests cold water.' On the one hand, in such responses the ambiguous NP was correctly interpreted as the object of the main verb. On the other hand, these paraphrases do not reflect understanding of the relative clause. Given these unpredicted difficulties with the coding of responses, we could not analyze the results from these conditions.

B. Appendix B: Analysis including ambiguous responses

In addition to the model reported in Section 3, we also fitted model using data of all participants, and including inconclusive results, to generate predictions for the individual average RP. In this analysis, we used data from all 92 participants, with three types of coded paraphrases: R (successful reanalysis), P (partial reanalysis), and inconclusive (word-by-word repetitions). For each participant, we calculated an RP average score over the four experimental sentences, counting R as 1 point, P as 0 points, and inconclusive as 0.5 points (so, e.g., a participant with four successful reanalyses had an average score of 1; an average score of 0.5 can reflect two successful and two partial reanalyses, or four inconclusive responses). We coded inconclusive responses with a value of 0.5 to reflect the intuition that a participant choosing word-by-word repetition is performing worse than a participant who chose to display their correct understanding of the sentence (e.g., by using a comma), but is still avoiding giving a wrong interpretation, perhaps noticing that it is illicit. We fitted a linear multiple regression model, modeling individual RP score as a function of two continuous variables, namely the participant's rs-EBR (with both linear and polynomial terms, given the literature on an inverse U-shaped relation between EBR and task performance) and the participant's log-transformed average critical RT, as well as their interaction: $lm(\text{average_reanalysis_performance} \sim RT \text{ of the critical region} \times \text{poly}(rs\text{-EBR}, 2), \text{data})$. RTs were logarithmically transformed and standardized, and rs-EBR values were standardized.

The results of this analysis are presented in Table B.1. The results are very similar to those of the main analysis, with a significant effect of Critical RT, a significant effect of both the linear and polynomial terms of rs-EBR, and no interaction between the two factors.

Figure B.1 presents the relation between rs-EBR and average RP for all 92 participants.

Table B.1. Results of regression model (N = 92)

Critical RT	rs-EBR	rs-EBR \times critical RT
$b = 0.1, t = 3.81, p = 0.0002$	Linear: $b = -0.70, t = -2.42, p = 0.017$ Polynomial: $b = -0.61, t = -2.0, p = 0.047$	$b = -0.57, t = -1.95, p = 0.053$ $b = 0.03, t = 0.08, p = 0.93$

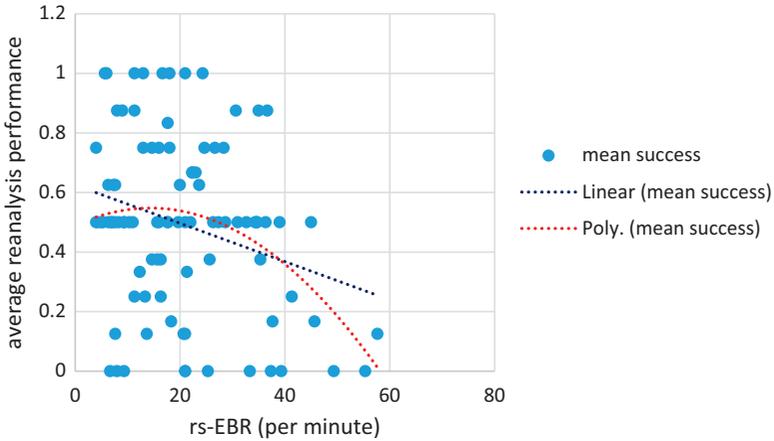


Fig. B1. Individual average reanalysis performance (RP) by rs-EBR ($N = 92$).

C. Appendix C: Models used in the analyses

Model 1. The main model of the study, predicting reanalysis performance based on blink rate and RTs, fitted for the 71 participants who had no ambiguous paraphrases:

$glm(\text{average_reanalysis_performance} \sim RT \text{ of the critical region} \times poly(rs\text{-EBR}, 2), \text{binomial}(\text{logit}), \text{weights} = n, \text{data})$

RTs were logarithmically-transformed and standardized, rs-EBR values were standardized.

Model 2. The GP effect analysis. A linear mixed model, modeling the log-transformed average RTs at the critical region, including the main verb and the sentence final region, as a function of condition (GP vs. baseline):

$lmer(RT \text{ of the critical region} \sim \text{condition} + (1 + \text{condition} | \text{subject}) + (1 | \text{item}), \text{data})$

RTs were logarithmically-transformed and standardized.

Model 3. The supplementary analysis, examining whether participants' ability to successfully reanalyze is correlated with their mean reading rate, measured as RT in filler sentences:

$glm(\text{average_reanalysis_performance} \sim RT \text{ of the critical region} \times poly(rs\text{-EBR}, 2) + \text{fillers RT}, \text{binomial}(\text{logit}), \text{weights} = n, \text{data}).$

RTs were logarithmically-transformed and standardized, rs-EBR values were standardized.