

Domain-General Conflict Monitoring Predicts Neural and Behavioral Indices of Linguistic Error Processing During Reading Comprehension

Trevor Brothers^{1, 2}, Margarita Zeitlin¹, Arim Choi Perrachione¹, Connie Choi¹, and Gina Kuperberg^{1, 2}

¹ Department of Psychology, Tufts University

² Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, Massachusetts, United States

The ability to detect and respond to linguistic errors is critical for successful reading comprehension, but these skills can vary considerably across readers. In the current study, healthy adults (age 18–35) read short discourse scenarios for comprehension while monitoring for the presence of semantic anomalies. Using a factor analytic approach, we examined if performance in nonlinguistic conflict monitoring tasks (Stroop, AX-CPT) would predict individual differences in neural and behavioral measures of linguistic error processing. Consistent with this hypothesis, domain-general conflict monitoring predicted both readers' end-of-trial acceptability judgments and the amplitude of a late neural response (the P600) evoked by linguistic anomalies. The influence on the P600 was nonlinear, suggesting that online neural responses to linguistic errors are influenced by both the effectiveness and efficiency of domain-general conflict monitoring. These relationships were also highly specific and remained after controlling for variability in working memory capacity and verbal knowledge. Finally, we found that domain-general conflict monitoring also predicted individual variability in measures of reading comprehension, and that this relationship was partially mediated by behavioral measures of linguistic error detection. These findings inform our understanding of the role of domain-general executive functions in reading comprehension, with potential implications for the diagnosis and treatment of language impairments.

Keywords: conflict monitoring, executive function, reading comprehension, individual differences, P600

Supplemental materials: <https://doi.org/10.1037/xge0001130.supp>

Comprehension is a goal-directed activity that allows us to infer an abstract message from strings of perceptual inputs. However, this goal is sometimes interrupted by the presence of linguistic errors. For example, in natural conversation, one in ten utterances contain some sort of speech error (Nakatani & Hirschberg, 1994), many of which are never corrected by the speaker (Levelt, 1983; Nooteboom, 1980). While error monitoring is a critical component of skilled comprehension, individuals can vary considerably in their ability to detect linguistic errors during online processing (Ehrlich et al., 1999; Garner, 1980; Hacker et al., 1998; Wagoner, 1983; Yudes et al., 2013). In the current study, we tested the hypothesis that linguistic error processing in healthy adult readers is affected by individual differences in domain-general *conflict monitoring*. Specifically, we asked whether the ability to monitor and resolve conflicts in nonlinguistic tasks can predict *neural* and *behavioral* measures of linguistic error processing, and whether this, in turn, has consequences for comprehension success.

Language errors are common in everyday comprehension and can arise from multiple sources. For example, overt errors occur frequently in natural speech and informal written communication, including mispronunciations, typos, and word substitutions. Even in carefully edited texts, *internal* processing errors can also arise on the part of the comprehender. For example, readers regularly encounter linguistic ambiguities that result in temporary misinterpretations (Altmann, 1998), while “slips of the eye” can lead readers to incorrectly recognize one word as another (Gregg & Inhoff, 2016; Kaufman & Obler, 1995). Because of these errors, comprehenders face a signal detection problem: was the current sentence understood as intended, or did it contain an error in need of reprocessing? To diagnose an error, readers must detect a *conflict* between the linguistic input and their current communication model—an internal model that encodes their high-level assumptions about the communicator and the broader linguistic environment (cf. Degen et al., 2015; Frank & Goodman, 2012; see Kuperberg et al., 2020, for recent discussion). Incoming words can conflict with a reader's communication model at multiple levels of representation, including orthography, syntax, and semantics. For example, most readers can rapidly infer that the sentence “He gave the candle the girl” contains a semantic error (Gibson et al., 2013) because the meaning of this sentence strongly conflicts with the reader's model of what is possible in the real world.

Once a conflict is detected, comprehenders can engage in additional compensatory behaviors like rereading, which allow them to

This article was published Online First November 29, 2021.

Funding was provided by the National Institute of Child Health and Human Development, Award ID: R01 HD08252.

Correspondence concerning this article should be addressed to Trevor Brothers, Department of Psychology, Tufts University, 490 Boston Avenue, Medford, MA 02155, United States. Email: trevor.brothers@gmail.com

reanalyze the input and potentially recover a text's intended meaning (Baker & Anderson, 1982; Carpenter & Daneman, 1981; Helder et al., 2016; Wagoner, 1983). For example, eye tracking studies have shown that highly implausible or syntactically unlicensed continuations are associated with longer fixation times (Kinnunen & Vauras, 1995; Traxler et al., 2000; Warren & McConnell, 2007; Warren et al., 2015) and more frequent regressions to earlier parts of the text (Ni et al., 1998; Rayner et al., 2004; Zabrocky & Ratner, 1989). Moreover, when readers are unable to make regressive saccades, their comprehension performance suffers, particularly in sentences that require semantic or syntactic reanalysis (Metzner et al., 2017; Schotter et al., 2014). This suggests that linguistic error detection and compensatory behaviors play a causal role in supporting text understanding. Indeed, in younger readers, the ability to consciously detect and respond to language errors is considered a central component of skilled reading (Baker & Anderson, 1982; Cain et al., 2004; Oakhill & Yuill 1996; Wagoner, 1983), and these *comprehension monitoring* skills are associated with improvements in reading comprehension across development (Oakhill & Cain, 2012).

In studies measuring event-related potentials (ERPs), linguistic anomalies have been associated with a late positive-going neural response—the *P600*—that is maximal from 600–1,000 ms over posterior electrode sites. While this response was initially associated with the processing of syntactic anomalies and ambiguities (Hagoort et al., 1993; Osterhout & Holcomb, 1992), later studies showed that *P600*s are elicited by a wide range of linguistic violations, including misspellings (Bulkes et al., 2020; Vissers et al., 2006) and semantic anomalies (Münte et al., 1998; Kuperberg et al., 2003; Szewczyk & Schriefers, 2011). For example, a robust *P600* is evoked by semantically anomalous sentence continuations (e.g., *Every morning for breakfast the eggs would *plant*), but not by mildly implausible continuations (e.g., *Every morning for breakfast the boys would plant*; Kuperberg et al., 2003, 2007; Paczynski & Kuperberg, 2012; van de Meerendonk et al., 2010). Simply detecting a semantic anomaly is not sufficient to generate this component; to produce a *P600*, readers must consciously perceive the input as a *comprehension error*, that is, they must be engaged in deep comprehension, and, as such, have established a communication model that conflicts with the input (Brothers et al., 2020; Kuperberg et al., 2020).¹ Once a conflict has been detected, readers can engage in additional second-pass mechanisms, such as *reanalysis*, to determine if a word was correctly perceived the first time around (see van de Meerendonk et al., 2009, for a discussion).²

The ability to monitor for conflict is thought to play a role in behavioral regulation in a wide range of nonlinguistic domains (Botvinick et al., 2001). In the cognitive control literature, the term *conflict monitoring* has been used to refer both to the automatic detection of conflict between competing responses (e.g., Yeung et al., 2004), as well as the ability to consciously detect and respond to conflicts between environmental inputs and an internal model of the current task (e.g., Yu et al., 2009). These “conflict signals” can then be used to regulate behavior, by inhibiting prepotent behavioral responses or selectively attending to relevant environmental inputs.³ For example, in the well-known Stroop task, the detection of conflict between well-learned aspects of the input (the meaning of the word “RED”) and a goal-relevant model that describes the task (identifying font colors), results in a reactive reallocation of attention to prevent incorrect responses.

To summarize thus far, during language comprehension, the detection of a conflict between the linguistic input and a comprehension-relevant communication model is thought to trigger the *P600*, which may reflect reanalysis or reallocation of attention to the prior linguistic input. Analogously, in nonlinguistic tasks, the detection of conflict between the input and a task-relevant internal model is thought to trigger a reallocation of attention to goal-relevant aspects of the input, to prevent behavioral errors. Despite these apparent similarities, it is unclear whether individual differences in domain-general conflict monitoring influence the online processing of linguistic errors in healthy adult readers.

Although there has been little systematic research into the relationship between the *P600* and domain-general conflict monitoring, some previous ERP studies have examined its links with a related construct—working memory capacity, which is often measured using “complex span” tasks (Daneman & Carpenter, 1980). The results of these studies are somewhat mixed. In one study (Nakano et al., 2010), high-span participants showed a robust *P600* response to semantically anomalous verbs while low-span participants showed an ERP effect of the opposite polarity—an *N400*, which has been linked to the difficulty of lexico-semantic retrieval (Kutas & Federmeier, 2011). A reading comprehension study by Kim et al. (2018) found a similar pattern: verbal (but not spatial) working memory capacity predicted the same trade-off relationship between *P600* and *N400* responses. On the other hand, several other studies report *no* significant relationships between working memory capacity and *P600* amplitudes, either in response to semantic anomalies (Kos et al., 2012; Ye & Zhou, 2008; Zheng & Lemhöfer, 2019), or syntactic anomalies (Tanner & van Hell, 2014; Tanner, 2019).

The results of these prior studies are somewhat difficult to interpret, both due to inconsistent findings and the multifactorial nature of working memory tasks. For example, performance in complex span tasks indexes a range of cognitive abilities, including

¹ Several internal and external factors can influence the probability of generating a *P600* response (Kuperberg, 2007). These include the nature of the experiment task (Payne et al., 2019), the length and constraint of the prior linguistic context (Brothers et al., 2020; Kuperberg et al., 2020; Nieuwland & Van Berkum, 2005), and, in impoverished contexts, the presence or absence of semantic attraction (e.g., Kuperberg, 2007; Kim & Osterhout, 2005).

² In the psycholinguistic literature, the resolution of linguistic errors has been discussed within multiple frameworks that are often specific to the type of error encountered. For example, the syntactic parsing literature has focused on the triggers for reanalysis in garden-path sentences (single-stage vs. two-stage accounts; see Traxler, 2014), as well as distinctions between the “revision” of syntactic ambiguities and the “repair” of syntactically ill-formed sentences (e.g., Gouvea et al., 2010; Kaan & Swaab, 2003; Vuong & Martin, 2014). The literature on semantic errors has also focused on different triggers for reanalysis (Kuperberg, 2007) and the potential resolution of anomalies through a noisy channel framework (Gibson et al., 2013). Here, we take a complementary and more general perspective, inspired by Kolk and colleagues, who proposed that these different forms of linguistic error processing are linked through a domain-general process of conflict monitoring (see Kolk & Chwilla, 2007; van de Meerendonk et al., 2009, for a discussion).

³ In control-demanding tasks, the presence of conflict can produce behavioral adjustments at multiple time scales, including the reactive reallocation of attention within the current trial (Gehring et al., 1993), adaptations to optimize performance on subsequent trials (Botvinick et al., 1999; Gratton et al., 1992), and longer-term learning that allows for the acquisition of new goals or task schemas (Botvinick, 2007).

memory storage, domain-specific verbal knowledge, and the inhibition of task irrelevant information (Bayliss et al., 2003; McCabe et al., 2010). Therefore, based on the previous literature, it is unclear which of these skills are most relevant for detecting linguistic errors (see Vuong & Martin, 2014, for a discussion).

The goal of the present study was to systematically examine the relationship between domain-general conflict monitoring and neural or behavioral measures of linguistic error processing in healthy adult readers. We recorded ERPs as participants read short discourse contexts for comprehension and monitored for the presence of semantic anomalies. To assess individual differences in linguistic error processing, we examined the P600 effect evoked by semantically anomalous (vs. plausible) words, as well as participants' end-of-trial plausibility judgments (d').

To examine the role of domain-general conflict monitoring, the same group of participants completed two conflict monitoring tasks: the Stroop task (MacLeod, 1991) and a fast-paced version of the AX Continuous Performance task (AX-CPT; Servan-Schreiber et al., 1996). These tasks were administered as part of a comprehensive neuropsychological battery, which also assessed individual differences in working memory capacity, verbal knowledge, and processing speed. We carried out a factor analysis to isolate latent "factor scores" for each of these cognitive domains, and these predictor variables were used to address two main theoretical questions.

First, does monitoring for linguistic errors during comprehension rely on domain-general conflict monitoring mechanisms? If so, then individual differences in domain-general conflict monitoring should predict both neural and behavioral responses to semantic errors during reading, even after controlling for variability in working memory, verbal knowledge, and processing speed. Here, we predicted that individuals with higher domain-general conflict monitoring scores would show higher behavioral accuracy and larger P600 amplitudes in response to semantic anomalies. We also considered an alternative "efficiency hypothesis" that would produce *dissociations* between behavioral and neural measures (Gray et al., 2005; Haier et al., 1992; Rypma et al., 2006). Specifically, individuals with better conflict monitoring abilities may resolve linguistic conflicts more efficiently, resulting in higher behavioral accuracy and smaller P600 responses overall.

Finally, we also examined the relationship between domain-general conflict monitoring and participants' ability to understand and answer questions about a text (*reading comprehension*). As noted earlier, monitoring and responding to linguistic errors is thought to be an important subcomponent of reading comprehension in both adult and developing readers (Cain et al., 2004; Garner, 1980; Wagoner, 1983). However, while it is well-established that working memory capacity and verbal abilities are important predictors of reading comprehension performance (Conway & Engle, 1996; Cromley et al., 2010; Daneman & Carpenter, 1980; Daneman & Merikle, 1996; Freed et al., 2017; Stanovich & Cunningham, 1992), it is less clear whether domain-general conflict monitoring plays an independent role (cf. Christopher et al., 2012; McVay & Kane, 2012). In the present study, we predicted that nonlinguistic conflict monitoring would account for additional variance in reading comprehension performance, and that this effect of domain-general conflict monitoring would be mediated by individual differences in linguistic error detection.

Method

Participants

The current study included data from 77 participants (43 female) between the ages of 18 and 35 ($M = 23.5$). This sample included 39 Tufts University undergraduates and 38 community-dwelling adults from the Boston metropolitan area, who were recruited using online advertisements. Participants were right-hand dominant, had no significant exposure to languages other than English before the age of five, and had no history of head injury or psychiatric or neurological diagnoses. Participants received course credit or were compensated for their participation. All protocols were approved by Tufts University Social, Behavioral, and Educational Research Institutional Review Board. All participants in this sample completed both the primary ERP study and a separate 2-hr session that included a range of behavioral individual differences measures (see below).

ERP Study

Experimental Design and Linguistic Stimuli

The experimental stimuli examined here consisted of 100 three-sentence discourse scenarios. In each scenario, the first two sentences introduced a coherent discourse context (*The lifeguards received a report of sharks right near the beach. Their primary concern was to prevent any incidents in the sea . . .*). The third sentence contained either an animate-constraining or inanimate-constraining verb followed by a critical noun that was either plausible (*Hence, they cautioned the trainees . . .*), or semantically anomalous (*Hence, they cautioned the drawer . . .*). To create the anomalous versions of each scenario, the same animate and inanimate critical nouns were counterbalanced across contexts.

Plausible and semantically anomalous nouns were matched in both lexical predictability ($<1\%$ cloze; Taylor, 1953) and mean cosine semantic similarity to the preceding context (latent semantic analysis, word-to-document similarity, $t < 1$; Landauer & Dumais, 1997). This is because both these factors are known to influence the amplitude of the N400 (Kutas & Federmeier, 2011). We then verified that the two conditions differed in plausibility by recruiting a separate group of participants from Amazon Mechanical Turk and asking them to rate each discourse scenario on a 7-point scale (7 = *makes perfect sense* to 1 = *makes no sense at all*). As expected, there were clear differences in plausibility ratings between the two conditions (plausible scenarios: $M = 5.5$, $SD = .9$, anomalous: $M = 1.9$, $SD = .6$). Additional information on the creation and norming of these stimuli are described in previous studies (Brothers et al., 2020; Kuperberg et al., 2020).

Plausible and anomalous scenarios were counterbalanced across lists, ensuring that participants saw each discourse context and critical word only once. To maximize statistical power, data were combined from three subgroups of participants who completed slightly different versions of the main ERP experiment (for full details, see online supplemental materials). These three versions used exactly the same presentation parameters and experimental items, and they all included equal proportions of plausible and semantically anomalous discourse scenarios. The primary differences were in the number of recording electrodes and in the

relative proportions of highly constraining and nonconstraining discourse scenarios. Due to differences in counterbalancing across samples, participants saw between 20 and 29 trials per condition within an experimental session.

Procedure

Participants sat in a sound-attenuated room, and discourse stimuli were presented on a computer monitor at a distance of 1.5 m. The first two sentences of each scenario were presented in their entirety, one sentence at a time, which participants read at their own pace. When participants pressed a button after the second sentence, the third critical sentence appeared one word at a time in the center of the screen (450 ms duration, 100 ms ISI). Participants were told to read the entire discourse carefully for comprehension. Following the sentence-final word, a question mark appeared after a 1,000 ms delay, and participants indicated via button press whether or not the preceding discourse “made sense.” On 20% of trials, participants also answered True/False comprehension questions that probed their understanding of the whole scenario.

EEG Preprocessing and Operationalization of ERP Components

Electroencephalogram (EEG) was recorded from a minimum of 32 scalp electrodes, arranged in a modified 10–20 system. Signals were digitized at 512 Hz with a bandpass filter of 0–104 Hz. Offline, EEG signals were referenced to the average of the right and left mastoids, and a .1–30 Hz bandpass filter was applied. The EEG was then segmented into epochs (–200 ms to 1,000 ms), time-locked to the onset of the critical noun. Independent component analysis was used to remove EEG artifact due to blinks, and any epochs with residual artifact were rejected (7% of trials). Artifact-free epochs were then averaged within-conditions for each participant.

Based on prior studies (Brothers et al., 2020; Kuperberg et al., 2020), we operationalized the P600 effect as the difference in amplitude between anomalous and plausible nouns from 600–1,000 ms over a cluster of posterior electrode sites (Pz, P3/4, Oz, O1/2). In addition to the P600 effect, we also examined differences in the N400 component, which has been shown to vary systematically with the amplitude of the P600 effect across participants (e.g., Kim et al., 2018). The N400 was operationalized as the average voltage from 300–500 ms over central-parietal electrode sites (CPz, CP1/2, Pz, P3/4).

Behavioral Assessment of Domain-General Cognitive Functions and Reading Comprehension

Behavioral Session: Task Procedures

In addition to the ERP experiment, participants also completed a 2-hr behavioral session on a separate day. This session included a battery of neuropsychological tasks assessing conflict monitoring, working memory, verbal knowledge, and processing speed, as well as two standardized reading comprehension assessments. Participants were tested individually in a quiet room by a trained experimenter. Nine of these tasks were administered by the experimenter, and five tasks were automated and presented on a desktop computer. The same task order was used for all participants. A short description of each task is provided below, along

with scoring methods and the dependent measures of interest (for additional information, see [online supplemental materials](#)).

Conflict Monitoring Tasks

To assess individual differences in conflict monitoring, participants completed a fast-paced version of the AX-CPT and a manual Stroop task. The AX-CPT task was first developed by Servan-Schreiber et al. (1996) to assess individual differences in proactive and reactive cognitive control. In this task, participants saw a letter cue (e.g., “A”) followed by a target (e.g., “X”). On the majority of trials (70%), participants saw a frequent cue-target pairing (“AX”), which requires a right-hand button response. On critical “AY” trials (10%), the “A” cue was followed by a different letter (e.g., “G”), and participants were required to withhold their default response and press a different key instead. The remaining 20% of trials served as control conditions, in which the target letter was preceded by a different cue (“BX” and “BY”; see Table 1).

Different versions of the AX-CPT have been developed to emphasize different aspects of cognitive control (Henderson et al., 2012). Here we used a fast-paced version of the task with short cue-target intervals (stimulus duration: 250 ms; SOA: 750 ms). With this version, by de-emphasizing the role of cue maintenance, we hoped to better dissociate variability in conflict monitoring and working memory. Participants completed a short practice session with feedback, followed by 150 cue-target pairings. *The two dependent measures for this task were AY response errors and the reaction time (RT) difference between critical and control trials (AY minus AX).*

In the manual Stroop task (MacLeod, 1991), participants identified the printed color of neutral letter strings (a string of X’s), congruent color words (“blue” in blue font), and incongruent color words (“blue” in red font). There were four possible button-press responses (red, black, green, and blue) and participants completed 28 trials in each condition. *The dependent measure for this task was percentage of errors in the incongruent condition.* We also calculated Stroop RT costs (*incongruent minus neutral*), but this measure was excluded from the factor analysis due to low reliability ($\alpha = .25$).

Table 1
Descriptive Statistics for Domain-General Conflict Monitoring Tasks

Variable	<i>M</i>	<i>SD</i>
AX-CPT		
AX Acc.	.965	.064
AY Acc.	.807	.177
BX Acc.	.934	.136
BY Acc.	.938	.124
AX RT (ms)	365	43
AY RT (ms)	476	50
BX RT (ms)	335	74
BY RT (ms)	341	72
Manual Stroop		
Congruent Acc.	.986	.027
Incongruent Acc.	.939	.051
Control Acc.	.979	.029
Congruent RT (ms)	968	261
Incongruent RT (ms)	1,176	318
Control RT (ms)	975	236

Note. Acc. = accuracy; AX-CPT = AX continuous performance task.

Working Memory Capacity Tasks

Four separate “complex span” tasks were administered that assessed participants’ ability to simultaneously manipulate and store information in working memory. They included automated versions of the Reading Span and Operation Span tasks (Unsworth et al., 2005), and experimenter administered versions of the Subtract-Two Span (Salthouse, 1988) and Listening Span tasks (Daneman & Carpenter, 1980). After a short practice session, participants completed three to five trials at each span length (see [online supplemental materials](#)). The dependent measure in each task was the total number of items recalled in all error-free sets (see Table 2).

Verbal Knowledge Tasks

Participants completed three tasks assessing their reading experience and verbal knowledge (Acheson et al., 2008; Uttl, 2002; Woodcock & Johnson, 1989). In the North American Adult Reading test (NAART) and Woodcock-Johnson Word Identification task (WORD ID), participants read lists of low frequency words aloud (*quadruped*, *leviathan*). In both tasks, the dependent measure was the number of correct pronunciations, which was scored by two independent raters using audio recordings. In the Author Recognition task (ART), participants selected known authors from a list of 65 famous authors and 65 nonfamous foils. The dependent measure was the number of correct identifications minus false alarms.

Processing Speed Tasks

To measure individual differences in verbal processing speed, three rapid naming tasks were administered: Rapid Automatized Naming (RAN), Rapid Alternating Stimulus (RAS), and the test of Word Reading Efficiency (TOWRE) Sight Word Efficiency task (Denckla & Rudel, 1976; Torgesen et al., 2012; Wolf, 1986). In these tasks, participants read printed lists of letters, letters and digits, or high-frequency words as quickly as possible without making mistakes. Because these tasks only provided a single outcome measure, measures of test-retest reliability were obtained from prior studies (Howe et al., 2006; Torgesen et al., 2012). The dependent measure for these tasks was time to completion, measured to the nearest second.

Reading Comprehension Assessments

In addition to these primary neuropsychological measures, participants also completed two reading comprehension tasks: (a) the comprehension portion of the Kauffman test of Educational Achievement (KTEA), which involves reading passages and answering multiple choice comprehension questions; and (b) the Woodcock Reading Mastery test (WRMT), which involves reading short passages and filling in a missing word (Singer et al., 2012; Woodcock, 1973). The dependent measure for these assessments was the total number of correct responses.

Preprocessing and Factor Analysis

The neuropsychological battery described above generated 13 performance measures in total. Before analysis, any scores more than 3 SDs above the group mean were replaced with this cutoff value (1.2% of scores). Average performance, excess skew and kurtosis, and the reliability for each dependent measure are shown in Table 2. As expected, these tasks showed acceptable to excellent levels of reliability ($\alpha = .62-.92$). A correlation matrix describing the relationships among these measures is presented in Table 3.

Using these scores, we conducted an exploratory factor analysis, which aimed to estimate the underlying latent constructs corresponding to our four cognitive abilities of interest. This analysis was carried out using SPSS 25. To maximize the independent variance of each latent variable, the factor analysis was fit using a Maximum Likelihood with a Varimax rotation. The total number of factors was selected using parallel analysis (Horn, 1965), and factor scores were calculated for each participant using the Bartlett regression method (Grice, 2001).

Consistent with the hypothesized structure of our neuropsychological battery, this exploratory factor analysis yielded a four-factor solution (see Table 4). Factor 1 explained 19% of the total variance and loaded strongly on measures of *working memory capacity*. Factor 2 loaded on measures of *processing speed* (16%), Factor 3 loaded on measures of *verbal knowledge* (13%), and Factor 4 loaded on measures of *conflict monitoring* (11%). Participant-

Table 2

Mean, Standard Deviation, Range, Skew, Kurtosis, and Reliability Values for Individual Difference Measures

Cognitive domain	Task	<i>M</i> (<i>SD</i>)	Range	Skew	Kurtosis	Reliability
Working memory capacity	Operation span	50.0 (17.7)	0–75	–1.0	1.2	.83 ^a
	Reading span	47.0 (18.7)	3–75	–0.5	–0.7	.82 ^a
	Listening span	29.8 (9.5)	13–55	0.1	–0.5	.89 ^a
	Subtract-2 span	128.0 (19.4)	70–169	–0.1	–0.1	.92 ^a
Verbal knowledge	NAART	40.4 (7.2)	19–53	–0.7	0.6	.89 ^a
	Word ID	14.0 (2.2)	7–17	–1.0	0.9	.75 ^a
	Author recognition	19.5 (8.6)	3–45	0.5	0.9	.83 ^b
Processing speed	RAN (s)	17.0 (3.4)	10–28	1.0	2.1	.91 ^c
	RAS (s)	18.7 (3.5)	11–30	0.5	0.7	.91 ^c
	TOWRE Sight (s)	48.6 (7.4)	72–35	0.6	0.4	.91 ^d
	Stroop accuracy	.934 (.051)	.79–1.0	–0.8	0.3	.62 ^a
Conflict monitoring*	AY - AX cost (ms)	110 (47)	9–250	0.4	0.2	.79 ^b
	AY accuracy	.807 (.177)	.40–1.0	–0.8	–0.3	.83 ^a

Note. NAART = North American Adult Reading test; Word ID = Woodcock-Johnson Word Identification task; RAN = Rapid Automatized Naming; RAS = Rapid Alternating Stimulus; TOWRE = test of Word Reading Efficiency.

^a Cronbach’s α . ^b Spearman-Brown split half. ^c Howe et al. (2006). ^d Torgesen et al. (2012).

* Stroop RT costs were excluded from the factor analysis due to low reliability ($\alpha = .25$).

Table 3*Correlation Matrix for Individual Differences Measures (N = 77)*

Task	1	2	3	4	5	6	7	8	9	10	11	12
1. O-SPAN	—											
2. R-SPAN	.65	—										
3. L-SPAN	.49	.49	—									
4. SUBT-2	.52	.54	.62	—								
5. ART	.12	.33	.22	.16	—							
6. Word_ID	.41	.43	.54	.50	.44	—						
7. NAART	.42	.50	.42	.52	.62	.78	—					
8. RAS	.30	.27	.21	.27	.00	.24	.25	—				
9. RAN	.34	.26	.29	.45	.08	.38	.35	.76	—			
10. TOWRE	.21	.32	.28	.26	.00	.27	.25	.54	.48	—		
11. Stroop _{acc}	.17	.22	.18	.26	−.02	.26	.24	−.09	−.01	.10	—	
12. AY _{acc}	.14	.20	.29	.14	.15	.46	.32	.17	.19	.24	.24	—
13. AY RT _{cost}	.05	.08	.12	.10	.05	.29	.17	−.12	.06	.11	.23	.56

Note. O-SPAN = operation span; R-SPAN = reading span; L-SPAN = listening span; SUBT-2 = subtract two span; ART = Author Recognition task; NAART = North American Adult Reading test; Word ID = Woodcock-Johnson Word Identification task; RAN = Rapid Automatized Naming; RAS = Rapid Alternating Stimulus; TOWRE = test of Word Reading Efficiency. Reaction times measures (RAS, RAN, TOWRE, and AY RT_{cost}) were reverse coded, with positive values indicating better performance. All bolded correlations ($r > .22$) were significant at $p < .05$.

specific factor scores were then used as predictor variables in a series of multiple regression analyses designed to test our a priori hypotheses, described below.

Hypothesis Testing

To test our primary hypotheses, we examined individual differences in two measures of linguistic error processing: (a) *behavioral detection of semantic errors*, as indexed by participants' acceptability judgments at the end of each scenario in the ERP experiment— d' scores; and (b) *neural semantic error processing*, as indexed by the magnitude of the P600 effect to anomalous versus plausible critical words. For each of these measures, we first examined the influence of

domain-general conflict monitoring using simple regression. We then carried out a multiple regression analysis, which included the conflict monitoring factor score and three other performance scores derived from our factor analysis. These multiple regression analyses enabled us to determine whether working memory, verbal knowledge, or processing speed predicted additional variability in our dependent measures, and whether any significant effects of conflict monitoring remained after controlling for these other factors. In these regression analyses, we included both linear and quadratic effects of each predictor to capture potential nonlinear contributions of each cognitive construct. We included these predictors because some prior studies have shown nonlinear relationships between executive functions and the magnitude of evoked neural responses (see Luna et al., 2010; Yarkoni & Braver, 2010, for reviews).

Finally, we performed a multiple regression analysis to ask whether domain-general conflict monitoring predicted variance in reading comprehension ability, beyond the effects of working memory and verbal knowledge. An additional mediation analysis tested whether any relationship between conflict monitoring and reading comprehension could be partially explained by individual differences in behavioral measures of linguistic error detection. In all multiple regression analyses, we standardized all predictors and dependent measures. This enabled us to compare the magnitude of effect sizes across analyses (*small*: $b = .1$; *medium*: $b = .3$; *large*: $b = .5$; Cohen, 1992; see [online supplemental materials](#)).

Results

ERP Study

Behavioral Detection of Semantic Errors: Acceptability Judgments

In the main ERP experiment, participants were able to categorize discourse scenarios as plausible ($M_{acc} = 87\%$, $SD = 11\%$) and anomalous ($M_{acc} = 90\%$, $SD = 10\%$). Although all participants performed above chance ($d' = 2.6$), there was also considerable

Table 4*Pattern Matrix for the Four-Factor Solution Obtained in an Exploratory Factor Analysis*

Task	Factor 1 (WMC)	Factor 2 (Speed)	Factor 3 (Verbal)	Factor 4 (Monitoring)
1. O-SPAN	.66	.23	.15	.01
2. R-SPAN	.63	.20	.24	.07
3. L-SPAN	.69	.16	.11	.21
4. SUBT-2	.75	.19	.21	.05
5. RAN	.11	.99	.03	−.10
6. RAS	.27	.74	.10	.02
7. TOWRE	.21	.54	.06	.16
8. ART	.12	−.02	.64	.05
9. WORD_ID	.43	.22	.58	.37
10. NAART	.37	.20	.89	.16
11. Stroop _{acc}	.27	−.09	.13	.28
12. AY _{acc}	.06	.24	.13	.81
13. AY RT _{cost}	.06	−.06	.05	.70

Note. O-SPAN = operation span; R-SPAN = reading span; L-SPAN = listening span; SUBT-2 = subtract two span; ART = Author Recognition task; NAART = North American Adult Reading test; Word ID = Woodcock-Johnson Word Identification task; RAN = Rapid Automatized Naming; RAS = Rapid Alternating Stimulus; TOWRE = test of Word Reading Efficiency; WMC = working memory capacity. Bolded values indicate the highest factor loading for each task.

variability in behavioral sensitivity across participants ($SD = .7$, range = 1.0 to 4.7, $\alpha = .69$).

Neural Index of Semantic Error Processing

Relative to plausible continuations, semantically anomalous words elicited a biphasic ERP response with larger (more negative) N400 amplitudes from 300–500 ms after word onset, and larger P600 amplitudes from 600–1,000 ms. As reported in previous studies (Kuperberg et al., 2020) these effects were both highly significant (*N400 effect*: $M = 1.1\mu V$, $SD = 2.5$, $t(76) = 3.96$, $p < .001$; *P600 effect*: $M = 2.2\mu V$, $SD = 3.3$, $t(76) = 5.94$, $p < .001$) and were maximal at posterior electrode sites.

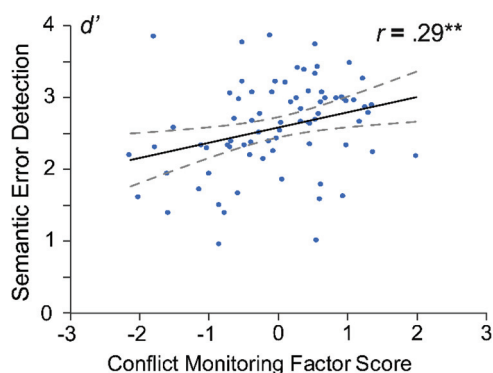
To determine the split-half reliability of this P600 anomaly effect, we calculated the amplitude of this ERP difference (*anomalous* vs. *plausible*) for each participant, separately for even and odd trials. After applying the Spearman-Brown prophecy formula, the P600 effect was moderately reliable across participants ($\rho = .61$).

Multiple Regression Analyses

A Linear Relationship Between Domain-General Conflict Monitoring and Behavioral Measure of Semantic Error Detection

Consistent with our main hypothesis, we observed a significant linear relationship between domain-general conflict monitoring and participants' ability to detect semantic anomalies, as indexed by their behavioral responses at the end of each scenario ($b = .29$, $t = 2.67$, $p = .009$; see Figure 1). This linear effect of conflict monitoring remained significant in a multiple regression analysis that included all four factor scores ($b = .27$, $t = 2.02$, $p = .047$). Participants with higher conflict monitoring abilities showed greater behavioral sensitivity to semantic errors, while working memory, verbal knowledge, and processing speed contributed no additional unique variance (see Table 5). We also replicated this effect of monitoring effect in an independent group of participants, using a different set of linguistic stimuli (see online supplemental materials).

Figure 1
The Linear Relationship Between Domain-General Conflict Monitoring and Semantic Error Detection During Comprehension



Note. Dotted lines represent 95% confidence intervals. $^{**}p < .01$. See the online article for the color version of this figure.

Table 5
Multiple Regression Analysis Predicting Semantic Error Detection (d')

Term	Factor score	β	t value	p -value
Linear	Monitoring	.27	2.02	.047
	WMC	.09	0.83	.41
	Verbal	.09	0.78	.44
	Speed	-.09	-0.73	.47
Quadratic	Monitoring ²	-.09	-0.65	.52
	WMC ²	-.16	-1.36	.18
	Verbal ²	-.21	-1.79	.08
	Speed ²	-.07	-0.49	.63

Note. $F(8, 68) = 2.28$; $p = .03$; $R^2 = .21$. WMC = working memory capacity. Significant predictors are indicated in bold, $p < .05$.

An Inverse U-Shaped Relationship Between Domain-General Conflict Monitoring and the P600 Effect

We next examined the effect of domain-general conflict monitoring on the magnitude of the P600 effect evoked by semantically anomalous (vs. plausible) critical words (Figure 2). We observed no significant linear effect, but instead saw a robust quadratic relationship between these two measures ($b = -.40$, $t = -3.76$, $p < .001$, for additional evidence see online supplemental materials). This quadratic relationship was due to relatively small P600 effects in individuals with low conflict monitoring abilities, a larger P600 effect in individuals in the middle of the scale, and a small P600 effect in participants with the strongest conflict monitoring abilities (see Figure 1). In a multiple regression analysis that included all four factors scores, the quadratic effect of domain-general conflict monitoring remained significant ($b = -.34$, $t = -2.52$, $p = .014$). Again, we saw no independent effects of working memory, verbal knowledge, or processing speed on the magnitude of the P600 effect (model $R^2 = .19$; see Table 6).

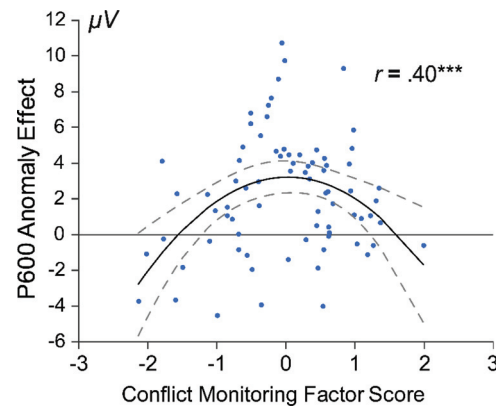
To confirm that this quadratic effect reflected an inverse U-shaped relationship, we used an interrupted regression analysis to test for the presence of a change in the sign of the regression coefficient (the two-line test; Simonsohn, 2018). Consistent with a U-shaped function, this analysis showed a positive relationship between domain-general conflict monitoring and the P600 effect over the lower half of the range ($b = 3.5$, $z = 3.46$, $p < .001$) and a negative relationship over the upper half of the range ($b = -2.4$, $z = -2.74$, $p = .006$).

To help visualize the influence of domain-general conflict monitoring on the magnitude of the P600 effect, we split participants into three groups (High, Medium, and Low) based on their conflict monitoring scores. As shown in Figure 3, individuals with intermediate scores showed the most robust P600 effects (4.0 μV), while smaller P600 effects were observed in both low conflict-monitoring (1.0 μV) and high conflict-monitoring participants (1.7 μV). As can be seen in Figure 3, ERP responses to plausible words were relatively constant across the three groups, and differences in the magnitude of the P600 effect were primarily driven by differential P600 responses to semantically anomalous words.

In addition, we also observed an apparent trade-off between the magnitude of the N400 and P600 effect across individuals. Consistent with prior studies (Kim et al., 2018; Nakano et al., 2010; Tanner, 2019; Tanner & van Hell, 2014), participants with larger P600 anomaly effects showed smaller N400 differences in the 300–500 ms time window (see Figure 3). Across all participants there was a

Figure 2

The Quadratic Relationship Between Domain-General Conflict Monitoring and the Magnitude of the P600 Effect to Semantic Anomalies Versus Plausible Control Words (600–1,000 ms)



Note. Dotted lines represent 95% confidence intervals. *** $p < .001$. See the online article for the color version of this figure.

significant negative correlation between the magnitude of the N400 and P600 effects to semantic anomalies, $r(75) = -.60, p < .001$.

To explore this relationship further, we conducted an additional multiple regression analysis, examining individual differences in the magnitude of the N400 effect (300–500 ms), see Table 7. As expected, this analysis also revealed a significant quadratic effect of conflict monitoring, which was opposite in sign to that observed in the P600 time-window ($b = .38, t = 2.90, p = .005$). Individuals with intermediate levels of conflict monitoring showed the smallest N400 differences ($-.4 \mu V$), and larger N400 effects were observed in both low-monitoring ($1.3 \mu V$) and high-monitoring participants ($2.3 \mu V$). As we note in the Discussion section, we attribute this reciprocal relationship to spatiotemporal overlap between the N400 and P600 components at the scalp surface.

Domain-General Conflict Monitoring and Behavioral Semantic Error Detection Predict Reading Comprehension

In our final set of analyses, we examined the effects of domain-general conflict monitoring on individual differences in reading comprehension. As expected, the two measures of reading comprehension were highly correlated ($r(75) = .57, p < .001$), and so we combined these scores into a single index.

A multiple-regression analysis revealed that both working memory ($b = .41, t = 4.75, p < .001$) and verbal knowledge ($b = .46, t = 5.26, p < .001$) were strongly associated with reading comprehension ability ($R^2 = .47$), replicating prior findings (Conway & Engle, 1996; Daneman & Merikle, 1996; Freed et al., 2017). We also observed an independent effect of domain-general conflict monitoring, with stronger conflict monitoring abilities predicting better reading comprehension ($b = .20, t = 2.33, p = .02$).

Finally, we asked whether this effect of domain-general conflict monitoring on reading comprehension could be partially explained by individual differences in linguistic error detection (d'). As shown in Figure 4, the indirect path (Conflict Monitoring $\rightarrow$

Semantic Error Detection $\rightarrow$ Comprehension) was significant (Sobel's test = 2.06, $p = .04$), and the direct effect of conflict monitoring was partially attenuated after accounting for variability in linguistic error detection (Control: $b = .14, t = 1.62, p = .11$). This suggests that the link between domain-general conflict monitoring and comprehension relies, to some extent, on differences in the ability to detect linguistic errors.

Discussion

We carried out a large individual differences study to examine the role of domain-general conflict monitoring in linguistic error processing. Participants read short discourse scenarios that sometimes contained semantic anomalies, and we examined variability across readers in both conscious error detection and the amplitude of the P600 response. Participants also completed a battery of tasks assessing domain-general conflict monitoring, working memory capacity, verbal knowledge, processing speed, and reading comprehension. We found that domain-general conflict monitoring was associated with behavioral error detection (with a linear relationship) and the magnitude of the P600 effect (with an inverse U-shaped relationship). We further showed that domain-general conflict monitoring predicted variability in reading comprehension ability—above and beyond the effects of working memory and verbal knowledge—and that this relationship could be partially explained by individual differences in semantic error detection. We discuss each of these findings in more detail below. We then discuss some open questions raised by our research as well as the theoretical and practical implications of our findings.

Domain-General Conflict Monitoring Predicts Both Behavioral and Neural Indices of Semantic Error Processing

The observed relationships between domain-general conflict monitoring and linguistic error processing were highly specific; that is, they could not be explained by variance in other cognitive abilities such as working memory capacity and verbal knowledge. We take this as evidence that *linguistic* error processing and *non-linguistic* conflict monitoring engage an overlapping set of cognitive processes, which involve detecting and responding to conflicts between environmental inputs and an internal mental model of the current task.

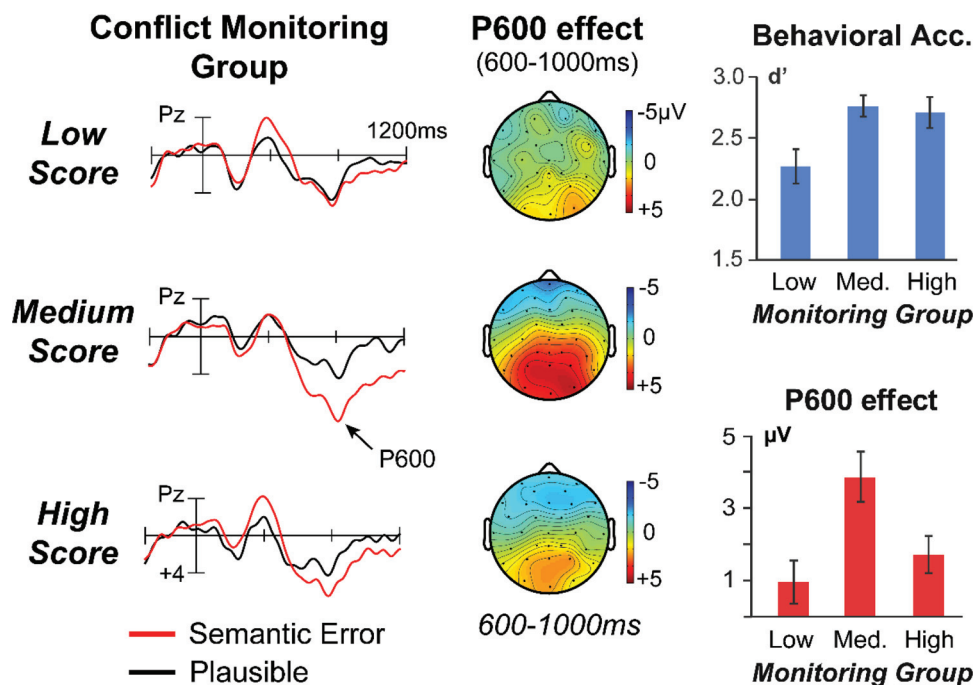
Obviously, the nature of these goal-relevant internal models differed during nonlinguistic conflict monitoring and monitoring for

Table 6
Multiple Regression Analysis Predicting P600 Effect Magnitudes

Term	Factor score	β	t value	p -value
Linear	Monitoring	.07	0.51	.61
	WMC	-.12	-1.11	.27
	Verbal	.01	0.07	.94
	Speed	.14	1.08	.29
Quadratic	Monitoring²	-.34	-2.52	.014
	WMC ²	-.05	-0.44	.66
	Verbal ²	.03	0.23	.82
	Speed ²	-.02	-0.15	.88

Note. $F(8, 68) = 2.04; p = .055; R^2 = .19$. WMC = working memory capacity. Significant predictors are indicated in bold, $p < .05$.

Figure 3
Differences in Error Processing Across Conflict Monitoring Groups



Note. Left: Evoked event-related potential (ERP) responses to plausible (black) and semantically anomalous (red) critical words, plotted separately for participants with Low, Medium, and High domain-general conflict monitoring abilities. Middle: Topographic plots showing the magnitude and distribution of the P600 effect in each group. Right: Mean behavioral sensitivity to semantic anomalies (top) and mean magnitude of the P600 effect (bottom) in each of the three groups. Error bars represent ± 1 SEM. See the online article for the color version of this figure.

linguistic errors during comprehension. For example, in the AX-CPT, participants were asked to respond whenever an “A” was followed by an “X.” To carry out this task efficiently and accurately, they likely engaged an internal model that represented their knowledge about the predictive relationships between cues and targets (Cohen et al., 1999; Yu et al., 2009). In contrast, during language processing, the goal was to *comprehend* the linguistic input (infer the communicator’s intended message). To achieve this goal, participants likely engaged an internal default “communication model” that represented their default linguistic and real-world knowledge (Kuperberg & Jaeger, 2016; Kuperberg et al., 2020). Critically, however, in both situations, participants sometimes encountered a bottom-up input that *conflicted* with this goal-relevant model (in the AX-CPT, an AY trial; in language comprehension, a semantic anomaly). We suggest that in both cases, the detection of this conflict led to a *disruption* of the default mode of processing, and triggered a reactive reallocation of attentional resources that enabled participants to achieve their goal. In the AX-CPT, this involved reallocating attention to the unexpected target letter, allowing for a shift to an alternative behavioral response (to avoid an error in action); in language comprehension, it involved reallocating attention to the critical word and prior context (*reanalysis*) in an attempt to reestablish coherence (avoiding a potential error in comprehension).

Although domain-general conflict monitoring predicted both behavioral and neural measures of semantic error processing, the

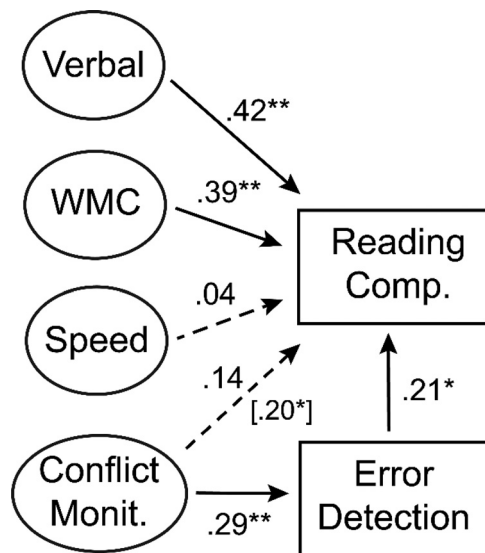
nature of these relationships differed: We observed a *linear* relationship with participants’ behavioral sensitivity to semantic errors (d'), and a *nonlinear* (inverse U-shaped) relationship with neural measures of error processing. Specifically, the P600 effect was largest in readers with moderate domain-general conflict monitoring ability and smallest in individuals with low or high conflict monitoring ability (see Figure 3). The same pattern of linear and nonlinear effects were also observed in a separate group of participants, using a different set of linguistic stimuli ($N = 37$; see [online supplemental materials](#)).

Table 7
Multiple Regression Analysis Predicting the Magnitude of the N400 Effect

Term	Factor score	β	t value	p -value
Linear	Monitoring	.21	1.64	.11
	WMC	.16	1.48	.14
	Verbal	.08	0.67	.50
	Speed	-.01	-0.10	.92
Quadratic	Monitoring²	.38	2.90	.005
	WMC ²	.16	1.40	.17
	Verbal ²	.09	0.76	.45
	Speed ²	.08	0.56	.58

Note. $F(8, 68) = 2.66$; $p = .013$; $R^2 = .24$. WMC = working memory capacity. Significant predictors are indicated in bold, $p < .05$.

Figure 4
A Path Diagram Representing the Effects of Verbal Knowledge, Working Memory Capacity (WMC), Processing Speed, and Domain-General Conflict Monitoring on Reading Comprehension



Note. The direct effect of domain-general conflict monitoring was partially attenuated after accounting for variability in semantic error detection (see text for explanation).

* $p < .05$. ** $p < .01$.

Although this type of dissociation between neural and behavioral measures is relatively novel in the ERP literature, similar dissociations have previously been reported in functional magnetic resonance imaging (fMRI) studies examining different aspects of executive function (see Luna et al., 2010; Manoach, 2003; Yarkoni & Braver, 2010, for reviews). For example, in developmental studies of the *antisaccade* task, hemodynamic responses within frontal control regions were largest in participants with intermediate behavioral performance (adolescents, ages 14–17), and smallest when antisaccade performance was either very poor (children, age 8–13) or close to ceiling (adults, age 18–30; Luna et al., 2001, 2010). Similar U-shaped response profiles have also been observed in studies that varied the difficulty of working memory demands. For example, in fMRI studies using the N-back task, large responses in control regions were observed at intermediate levels of cognitive load, while reduced activity was observed when cognitive load was minimal or when load was very high and behavioral performance began to break down (Callicott et al., 1999; Ciesielski et al., 2006; Mattay et al., 2006; see also Vogel & Machizawa, 2004).

In these previous neuroimaging studies, inverse U-shaped responses have been interpreted as reflecting a trade-off between the *effectiveness* and *efficiency* of executive control mechanisms (Gray et al., 2005; Haier et al., 1992; Larson et al., 1995; Rypma et al., 2006). For example, when participants are unable to carry out a task correctly, they may fail to engage control mechanisms at all. At moderate levels of task difficulty, accurate task performance engages maximum executive resources, resulting in the largest neural activity. Finally, when participants are highly skilled in a task, they are able to carry it out with increased efficiency, resulting in reduced neural responses.

In the present study, we suggest that participants with the worst domain-general conflict monitoring performance had the most difficulty detecting semantic errors, as evidenced by their poor behavioral performance and relatively small P600 effects (see Batterink & Neville, 2013; Sanford et al., 2011, for evidence that only detected anomalies elicit P600 responses). In contrast, participants with intermediate domain-general conflict monitoring were better able to detect the presence of semantic anomalies and may have been more likely to engage neural resources to reanalyze the linguistic input, resulting in a robust P600 response. Finally, for participants with high conflict monitoring abilities, we believe they exhibited high accuracy and small P600 effects because their neural processing was maximally efficient. For example, these high-performing participants may have been able to efficiently categorize semantic errors as anomalous without engaging in extensive reprocessing of the critical word or the prior context. Alternatively, they may need fewer neural resources to reanalyze the input, compared with other readers.

Note, the account outlined above assumes that our behavioral measure (d') primarily reflected comprehenders' abilities to consciously *detect* the presence of linguistic errors, while the magnitude of the P600 effect was sensitive, not just to error detection (Batterink & Neville, 2013; Sanford et al., 2011) but also additional processing stages, including a reanalysis of the anomalous input and the prior context (see Brothers et al., 2020; Metzner et al., 2017).

N400-P600 Tradeoffs

In addition to predicting the magnitude of the P600 anomaly effect, we also observed an effect of domain-general conflict monitoring on the magnitude of the N400 effect, which is thought to reflect the difficulty of lexico-semantic access/retrieval (Kutas & Federmeier, 2011). This N400 conflict monitoring effect was exactly opposite to the relationship observed for the P600, with smaller N400 differences in readers with moderate conflict monitoring ability, and larger N400 differences for individuals with either low or high conflict monitoring ability. Reciprocal relationships between N400 and P600 amplitudes have been observed in previous ERP studies examining semantic or syntactic anomalies (Kim et al., 2018; Kos, van den Brink & Hagoort, 2012; Nakano et al., 2010; Tanner, 2019; Tanner & van Hell, 2014). It has been suggested by some researchers that this reciprocal relationship reflects a trade-off in processing strategies as readers attempt to resolve competing linguistic constraints (Kim et al., 2018). According to this trade-off account, "P600 dominant" individuals are more likely to resolve semantic anomalies through structural reanalysis (*The meals were devouring* → *The meals were devoured*), while "N400 dominant" individuals are more likely to attempt to retrieve the semantic features of anomalous critical words (Kim & Osterhout, 2005).

Given that the semantic errors in the present experiment had no plausible syntactic edit (*Hence, they cautioned the drawer* ... → ??), we believe that an alternative explanation is more likely. Specifically, we suggest that the negative correlation between N400 and P600 amplitudes simply reflects spatiotemporal overlap of ERP components with opposite polarities (for discussion, see Brouwer & Crocker, 2017; Kuperberg et al., 2007; Tanner et al., 2018). EEG responses always reflect a combination of multiple

neural generators, and ERPs with opposite polarities are known to “cancel out” at the scalp surface. Because N400 and P600 responses often have similar scalp distributions and overlapping time courses, any increase in the amplitude of the P600 will result in a decrease the magnitude of the N400 (and vice-versa). We suggest that *all three* monitoring groups had greater difficulty retrieving the meaning of semantically anomalous words, resulting in an N400 effect. However, differences in conflict monitoring affected their ability to consciously detect and reanalyze these anomalies. This resulted in greater P600 responses for some participants, leading to greater amounts of component cancellation in the N400 time-window. Therefore, rather than reflecting a *cognitive* trade-off between semantic and structural processing, we believe this negative correlation simply reflected the cancellation of two independent neural generators.

No Effect of Working Memory Capacity or Verbal Knowledge on Semantic Error Processing

Unlike domain-general conflict monitoring, neither working memory capacity nor verbal knowledge were predictive of individual differences in linguistic error processing. Both these predictors showed high levels of internal validity and were strongly associated with individual differences in reading comprehension (see below). Therefore, it is unlikely that these null results reflected issues of measurement error or construct validity. Instead, they suggest that individual variability in working memory and verbal knowledge do not contribute substantially to linguistic error processing in skilled adult readers.

The absence of a relationship between the P600 effect and working memory capacity is consistent with some but not all previous findings in the literature. As noted in the Introduction, some studies have reported positive correlations between working memory and the magnitude of the P600 (Kim et al., 2018; Nakano et al., 2010), while others have found no significant relationship (Kos et al., 2012; Tanner, 2019; Zheng & Lemhöfer, 2019). In the present study, despite the inclusion of multiple span measures and a relatively large sample, we observed no significant correlations, either at the level of individual tasks (Subtract-2: $r = .01$, LSPAN: $r = .06$, RSPAN: $r = -.02$, OSPAN: $r = .03$), or for our combined working memory factor score, ($r(75) = -.06$, $p = .58$). One possibility is that this null effect can be explained by differences in the participants or linguistic stimuli across studies (e.g., discourse scenarios vs. single sentences). Another possibility is that the relationship between working memory capacity and the P600 is either nonexistent, or so small that it is difficult to detect reliably in a single experiment. Ultimately, a combination of preregistered ERP studies and a systematic meta-analysis of the literature may be needed to definitively resolve this issue.

At face value, the null relationship between verbal knowledge and linguistic error processing appears to contradict some prior studies examining second-language (L2) learners. In these studies, low-proficiency L2 learners often show absent or reduced P600 effects compared with native speakers (Osterhout et al., 2006; Zheng & Lemhöfer, 2019). Moreover, the magnitude of the P600 in these groups correlates with both error detection rates and measures of L2 proficiency (Tanner et al., 2013; Zheng & Lemhöfer, 2019). Critically, however, unlike most native English speakers, lower-proficiency L2 learners are likely to lack some of the core

semantic and syntactic knowledge necessary to detect linguistic anomalies.⁴ In light of these results, our current findings suggest that, as readers reach native-like proficiency, subtle differences in verbal knowledge become less important. Instead, in this sample of native speakers, most of the variability in error monitoring performance depended on differences in nonlinguistic conflict monitoring, which may regulate the successful *application* of stored linguistic knowledge during real-time comprehension.

Domain-General Conflict Monitoring Predicts Measures of Reading Comprehension

In addition to examining its relationship with linguistic error processing, we were also interested in whether domain-general conflict monitoring predicted individual differences in comprehension ability, as indexed by standardized measures of reading comprehension. From previous studies, it is clear that comprehension abilities in adult readers vary as a function of working memory capacity (Conway & Engle, 1996; Daneman & Carpenter, 1980; Daneman & Merikle, 1996) and verbal knowledge (Cromley et al., 2010; Freed et al., 2017; Stanovich & Cunningham, 1992). In the current study, we replicated these findings: working memory capacity and verbal knowledge uniquely accounted for 16% and 20% of the variance in reading comprehension performance.

Even after controlling for differences in working memory, verbal knowledge, and processing speed, we found that approximately 4% of the additional variance in reading comprehension ability could be explained by individual differences in domain-general conflict monitoring. Additional analyses suggested that this conflict monitoring effect was partially mediated by participants' ability to detect linguistic errors. To explain these relationships, we hypothesize that skilled readers continually monitor for potential errors in comprehension to maintain coherence (van de Meerendonk et al., 2009). If a reader fails to detect these conflicts, or is unable to resolve them efficiently, they may continue with an incorrect or internally contradictory interpretation of a text's meaning (Oakhill & Yuill 1996), which may have negative, downstream consequences for comprehension. On this account, *individuals with poor domain-general conflict monitoring abilities are poorer comprehenders because they are less able to detect and resolve to their own processing errors.*

Notably, domain-general conflict monitoring accounted for a smaller proportion of variance in comprehension performance (4%) compared with the effects of working memory capacity (16%) and verbal knowledge (20%). This suggests that conflict monitoring mechanisms may play a more specialized role in reading comprehension, intervening relatively infrequently to resolve processing errors that would otherwise disrupt comprehension. This may explain why, in some fMRI studies of “naturalistic” language comprehension, brain regions associated with conflict monitoring are infrequently reported (Fedorenko, 2014), particularly

⁴ Consistent with this suggestion, in our multiple regression analysis, we observed a marginally significant quadratic effect of verbal knowledge on semantic error detection (Verbal 2: $b = -.10$, $t = 1.79$, $p = 0.08$). Specifically, in participants with very low verbal knowledge scores (the bottom tertile), error detection was less accurate ($d' = 2.3$, $SD = 0.5$) than participants with higher verbal knowledge ($d' = 2.7$, $SD = 0.7$).

when the hemodynamic response is not time-locked to the onset of comprehension problems.

Predictions and Open Questions

This domain-general conflict monitoring account raises some important questions and generates a number of testable predictions. First, do these results generalize to other types of linguistic error processing? In the current experiment, we examined individual differences in *semantic* error processing using local verb-argument mismatches. In future studies it will be important to determine whether domain-general conflict monitoring also influences sensitivity to other types of linguistic errors that are known to generate P600 effects (e.g., Osterhout & Holcomb, 1992; Vissers et al., 2006). For example, readers with poor domain-general monitoring may be more likely to overlook misspellings or grammatical errors as they read, and they may experience difficulties when resolving lexical or syntactic ambiguities (Engelhardt et al., 2017; Vuong & Martin, 2014). Similarly, future studies should also examine whether different types of linguistic anomalies also show a U-shaped relationship between conflict monitoring and the magnitude of the P600 response.

It will also be important to determine whether similar domain-general performance monitoring mechanisms are involved in the detection and resolution of more *global* discourse comprehension errors. During discourse comprehension, readers must encode abstract information that unfolds across multiple sentences, including thematic content, causal relationships, and character motivations (Baker & Anderson, 1982; Kinnunen & Vauras, 1995). Similar to local semantic anomalies, the detection of global conflicts can also indicate the onset of comprehension problems (e.g., *a vegetarian protagonist eating a hamburger*), which may prompt the reader to revise or reinterpret prior contextual information (Albrecht & O'Brien, 1993; Braasch et al., 2012; Hakala & O'Brien, 1995). Although related, we believe that the detection of global discourse anomalies may play an even more important role in successful text comprehension (Oakhill et al., 2005), potentially relying on both domain-general conflict monitoring and working memory resources.

Given our current findings, it is also important to consider the relationship between error processing in *comprehension* and the error monitoring processes involved in language *production*. It is well established that speakers continually monitor their own production to prevent or correct unintended speech errors. Many aspects of error monitoring in language production are relatively automatic, involving the detection and resolution of internal representational conflicts within the production system (Nozari et al., 2011), which may be linked to certain domain-general executive functions (e.g., Altmann et al., 2001; Gollan et al., 2020). Here we suggest that, even more relevant to the comprehension error monitoring construct examined in the present study, are those aspects of error monitoring in language production that are less automatic, involving the conscious detection of errors and the initiation of downstream repairs (e.g., “*turn left . . . uh, right*”). Traditional monitoring accounts of language production have argued that these latter types of error monitoring processes are carried out by the producer's own comprehension system (monitoring through a perceptual loop, Levelt, 1983; see Gauvin & Hartsuiker, 2020, for a recent discussion). We suggest that linguistic error detection

during *both* language comprehension and production may rely on a common set of computational mechanisms that are also engaged in aspects of domain-general conflict monitoring (Yu et al., 2009). In future studies it will be important to test this hypothesis more directly by examining the neural and behavioral correlates of error monitoring during both language comprehension and production.

Neurobiology of Conflict Monitoring

The error monitoring account discussed in this study also makes predictions regarding the relationship between linguistic error processing and neurobiological measures of conflict monitoring. In previous studies, conflict-driven attention shifts have been associated with the phasic release of norepinephrine (NE) in the locus coeruleus (Aston-Jones & Cohen, 2005), which is thought to serve as an “orienting signal” to interrupt the default task state (Bouret & Sara, 2005; Dayan & Yu, 2006; Yu & Dayan, 2005). The phasic release of NE has also been linked with positive-going ERP responses like the P300 (Nieuwenhuis et al., 2005; Vazey et al., 2018), which is thought to be functionally related to the P600 response (Coulson et al., 1998; Osterhout et al., 2012; Sassenhagen et al., 2014; Sassenhagen & Fiebach, 2019). Also of relevance is a domain-general ERP component known as the “error positivity” or *Pe* (Falkenstein et al., 2000), which has been associated with the conscious, “metacognitive” recognition of errors (Nieuwenhuis et al., 2001) and compensatory processes, such as additional information seeking, in the service of optimizing performance (Desender et al., 2019). It will be important for future studies to investigate whether linguistic error detection is associated with physiological markers of phasic NE release, such as pupil dilation (Sara, 2009), and whether P600 amplitudes are correlated with other positive-going ERP responses elicited in classic error monitoring tasks.

In fMRI studies, conflict-inducing tasks such as the Stroop and AX-CPT have been shown to activate a fronto-parietal network of “cognitive control regions,” including the anterior cingulate and lateral prefrontal cortex (Niendam et al., 2012; van Veen & Carter, 2002). If linguistic conflict detection relies on the same neurocognitive mechanisms, then a similar set of brain regions should be recruited when comprehenders detect linguistic errors. Indeed, in fMRI studies, semantic anomalies and nonlinguistic conflicts (*incongruent Stroop trials*) have been shown to activate overlapping regions of left inferior frontal cortex (van de Meerendonk et al., 2013; Ye & Zhou, 2009). Similar coactivations within this region have also been observed for sentences with syntactic ambiguities (Hsu et al., 2017), which readers may initially interpret as syntactic errors (see Novick et al., 2010; Nozari & Thompson-Schill, 2016, for reviews).

We should note, however, that this does not imply that these frontal regions contribute directly to the posterior P600 effect observed at the scalp surface (see van de Meerendonk et al., 2013, for a discussion). For example, frontal activity may reflect the initial *detection* of conflict, which then triggers additional compensatory mechanisms that are reflected in the P600 itself. Consistent with this idea, in a recent multimodal neuroimaging study (Wang et al., 2021), semantic anomalies produced robust activations within left inferior frontal cortex and anterior cingulate, from 300–500 ms after word onset. Critically, this activity was followed by a late reactivation of fusiform cortex (from 600–800 ms) that

more closely matched the timing of the P600 response. This late fusiform response may have reflected an orthographic reanalysis of the bottom-up input following an error.

Practical Applications

Our findings also have potential applications for the diagnosis and treatment of reading comprehension impairments. Reading comprehension problems are one of the most frequently observed deficits in children with learning disabilities (Lyon, 1995; Vaughn et al., 2002). In addition to problems with vocabulary and phonological awareness, poor readers often show impairments in their ability to actively monitor their comprehension as they read. For example, poor readers are less accurate at detecting linguistic anomalies (Garner, 1980; Hacker et al., 1998; Oakhill & Cain, 2012; Rubman & Salatas Waters, 2000), and, when comprehension errors are detected, these readers are less likely to engage in useful compensatory strategies such as rereading (Ehrlich et al., 1999; Hacker et al., 1998). Given our findings in adults, it is possible that children with poor domain-general conflict monitoring may be predisposed to develop reading comprehension problems (Cutting et al., 2009). If this is the case, then early screening for conflict monitoring deficits could be useful tool for identifying at-risk readers, who may benefit from additional instruction or targeted interventions (Gersten et al., 2001; Suggate, 2016).

Finally, a number of neuropsychiatric disorders that affect language processing have been linked to deficits in domain-general conflict monitoring, including schizophrenia (Boudewyn et al., 2012; Kerns et al., 2005; Kuperberg, 2010; Lesh et al., 2011) and autism spectrum disorder (Agam et al., 2010; Solomon et al., 2008; South et al., 2010). Consistent with our current findings, individuals with these neurodevelopmental disorders also show marked impairments in linguistic error detection and the magnitude of the P600 response (Koolen et al., 2013; Kuperberg et al., 1998; Kuperberg et al., 2006). An important goal for future research will be to determine whether similar abnormalities in linguistic error processing are observed across diagnostic boundaries (Insel et al., 2010) and whether they can be explained by domain-general conflict monitoring deficits.

References

- Acheson, D. J., Wellu, J. B., & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods*, 40(1), 278–289. <https://doi.org/10.3758/BRM.40.1.278>
- Agam, Y., Joseph, R. M., Barton, J. J. S., & Manoach, D. S. (2010). Reduced cognitive control of response inhibition by the anterior cingulate cortex in autism spectrum disorders. *NeuroImage*, 52(1), 336–347. <https://doi.org/10.1016/j.neuroimage.2010.04.010>
- Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(5), 1061–1070. <https://doi.org/10.1037/0278-7393.19.5.1061>
- Altmann, G. T. M. (1998). Ambiguity in sentence processing. *Trends in Cognitive Sciences*, 2(4), 146–152. [https://doi.org/10.1016/S1364-6613\(98\)01153-X](https://doi.org/10.1016/S1364-6613(98)01153-X)
- Altmann, L. J. P., Kempler, D., & Andersen, E. S. (2001). Speech errors in Alzheimer's disease: Reevaluating morphosyntactic preservation. *Journal of Speech, Language, and Hearing Research*, 44(5), 1069–1082. [https://doi.org/10.1044/1092-4388\(2001/085\)](https://doi.org/10.1044/1092-4388(2001/085))
- Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, 28(1), 403–450. <https://doi.org/10.1146/annurev.neuro.28.061604.135709>
- Baker, L., & Anderson, R. I. (1982). Effects of inconsistent information on text processing: Evidence for comprehension monitoring. *Reading Research Quarterly*, 17(2), 281–294. <https://doi.org/10.2307/747487>
- Batterink, L., & Neville, H. J. (2013). The human brain processes syntax in the absence of conscious awareness. *The Journal of Neuroscience*, 33(19), 8528–8533. <https://doi.org/10.1523/JNEUROSCI.0618-13.2013>
- Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General*, 132(1), 71–92. <https://doi.org/10.1037/0096-3445.132.1.71>
- Botvinick, M. M. (2007). Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function. *Cognitive, Affective & Behavioral Neuroscience*, 7(4), 356–366. <https://doi.org/10.3758/CABN.7.4.356>
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3), 624–652. <https://doi.org/10.1037/0033-295X.108.3.624>
- Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature*, 402(6758), 179–181. <https://doi.org/10.1038/46035>
- Boudewyn, M. A., Carter, C. S., & Swaab, T. Y. (2012). Cognitive control and discourse comprehension in schizophrenia. *Schizophrenia Research and Treatment*, 2012, 484502. <https://doi.org/10.1155/2012/484502>
- Bouret, S., & Sara, S. J. (2005). Network reset: A simplified overarching theory of locus coeruleus noradrenergic function. *Trends in Neurosciences*, 28(11), 574–582. <https://doi.org/10.1016/j.tins.2005.09.002>
- Braasch, J. L. G., Rouet, J.-F., Vibert, N., & Britt, M. A. (2012). Readers' use of source information in text comprehension. *Memory & Cognition*, 40(3), 450–465. <https://doi.org/10.3758/s13421-011-0160-6>
- Brothers, T., Wlotko, E. W., Warnke, L., & Kuperberg, G. R. (2020). Going the extra mile: Effects of discourse context on two late positivities during language comprehension. *Neurobiology of Language*, 1(1), 135–160. https://doi.org/10.1162/nol_a_00006
- Brouwer, H., & Crocker, M. W. (2017). On the proper treatment of the N400 and P600 in language comprehension. *Frontiers in Psychology*, 8, 1327. <https://doi.org/10.3389/fpsyg.2017.01327>
- Bulkes, N. Z., Christianson, K., & Tanner, D. (2020). Semantic constraint, reading control, and the granularity of form-based expectations during semantic processing: Evidence from ERPs. *Neuropsychologia*, 137, 107294. <https://doi.org/10.1016/j.neuropsychologia.2019.107294>
- Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, 96(1), 31–42. <https://doi.org/10.1037/0022-0663.96.1.31>
- Callicott, J. H., Mattay, V. S., Bertolino, A., Finn, K., Coppola, R., Frank, J. A., Goldberg, T. E., & Weinberger, D. R. (1999). Physiological characteristics of capacity constraints in working memory as revealed by functional MRI. *Cerebral Cortex*, 9(1), 20–26. <https://doi.org/10.1093/cercor/9.1.20>
- Carpenter, P. A., & Daneman, M. (1981). Lexical retrieval and error recovery in reading: A model based on eye fixations. *Journal of Verbal Learning and Verbal Behavior*, 20(2), 137–160. [https://doi.org/10.1016/S0022-5371\(81\)90357-1](https://doi.org/10.1016/S0022-5371(81)90357-1)
- Christopher, M. E., Miyake, A., Keenan, J. M., Pennington, B., DeFries, J. C., Wadsworth, S. J., Willcutt, E., & Olson, R. K. (2012). Predicting word reading and comprehension with executive function and speed measures across development: A latent variable analysis. *Journal of*

- Experimental Psychology: General*, 141(3), 470–488. <https://doi.org/10.1037/a0027375>
- Ciesielski, K. T., Lesnik, P. G., Savoy, R. L., Grant, E. P., & Ahlfors, S. P. (2006). Developmental neural networks in children performing a Categorical N-Back Task. *NeuroImage*, 33(3), 980–990. <https://doi.org/10.1016/j.neuroimage.2006.07.028>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Cohen, J. D., Barch, D. M., Carter, C., & Servan-Schreiber, D. (1999). Context-processing deficits in schizophrenia: Converging evidence from three theoretically motivated cognitive tasks. *Journal of Abnormal Psychology*, 108(1), 120–133. <https://doi.org/10.1037/0021-843X.108.1.120>
- Conway, A. R., & Engle, R. W. (1996). Individual differences in working memory capacity: More evidence for a general capacity theory. *Memory*, 4(6), 577–590. <https://doi.org/10.1080/741940997>
- Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, 13(1), 21–58. <https://doi.org/10.1080/016909698386582>
- Cromley, J. G., Snyder-Hogan, L. E., & Luciw-Dubas, U. A. (2010). Reading comprehension of scientific text: A domain-specific test of the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*, 102(3), 687–700. <https://doi.org/10.1037/a0019452>
- Cutting, L. E., Materek, A., Cole, C. A. S., Levine, T. M., & Mahone, E. M. (2009). Effects of fluency, oral language, and executive function on reading comprehension performance. *Annals of Dyslexia*, 59(1), 34–54. <https://doi.org/10.1007/s11881-009-0022-0>
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450–466. [https://doi.org/10.1016/S0022-5371\(80\)90312-6](https://doi.org/10.1016/S0022-5371(80)90312-6)
- Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, 3(4), 422–433. <https://doi.org/10.3758/BF03214546>
- Dayan, P., & Yu, A. J. (2006). Phasic norepinephrine: A neural interrupt signal for unexpected events. *Network*, 17(4), 335–350. <https://doi.org/10.1080/09548980601004024>
- Degen, J., Tessler, M. H., & Goodman, N. D. (2015). Wonky worlds: Listeners revise world knowledge when utterances are odd. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 548–553). Cognitive Science Society.
- Denckla, M. B., & Rudel, R. G. (1976). Rapid “automatized” naming (R.A.N): Dyslexia differentiated from other learning disabilities. *Neuropsychologia*, 14(4), 471–479. [https://doi.org/10.1016/0028-3932\(76\)90075-0](https://doi.org/10.1016/0028-3932(76)90075-0)
- Desender, K., Murphy, P., Boldt, A., Verguts, T., & Yeung, N. (2019). A postdecisional neural marker of confidence predicts information-seeking in decision-making. *The Journal of Neuroscience*, 39(17), 3309–3319. <https://doi.org/10.1523/JNEUROSCI.2620-18.2019>
- Ehrlich, M.-F., Remond, M., & Tardieu, H. (1999). Processing of anaphoric devices in young skilled and less skilled comprehenders: Differences in metacognitive monitoring. *Reading and Writing*, 11(1), 29–63. <https://doi.org/10.1023/A:1007996502372>
- 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>
- Falkenstein, M., Hootmann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: A tutorial. *Biological Psychology*, 51(2–3), 87–107. [https://doi.org/10.1016/S0301-0511\(99\)00031-9](https://doi.org/10.1016/S0301-0511(99)00031-9)
- Fedorenko, E. (2014). The role of domain-general cognitive control in language comprehension. *Frontiers in Psychology*, 5, 335. <https://doi.org/10.3389/fpsyg.2014.00335>
- Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336(6084), 998. <https://doi.org/10.1126/science.1218633>
- 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>
- Garner, R. (1980). Monitoring of understanding: An investigation of good and poor readers’ awareness of induced miscomprehension of text. *Journal of Reading Behavior*, 12(1), 55–63. <https://doi.org/10.1080/10862968009547352>
- Gauvin, H. S., & Hartsuiker, R. J. (2020). Towards a new model of verbal monitoring. *Journal of Cognition*, 3(1), 17. <https://doi.org/10.5334/joc.81>
- Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science*, 4(6), 385–390. <https://doi.org/10.1111/j.1467-9280.1993.tb00586.x>
- Gersten, R., Fuchs, L. S., Williams, J. P., & Baker, S. (2001). Teaching reading comprehension strategies to students with learning disabilities: A review of research. *Review of Educational Research*, 71(2), 279–320. <https://doi.org/10.3102/00346543071002279>
- Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences of the United States of America*, 110(20), 8051–8056. <https://doi.org/10.1073/pnas.1216438110>
- Gollan, T. H., Li, C., Stasenko, A., & Salmon, D. P. (2020). Intact reversed language-dominance but exaggerated cognate effects in reading aloud of language switches in bilingual Alzheimer’s disease. *Neuropsychology*, 34(1), 88–106. <https://doi.org/10.1037/neu0000592>
- Gollan, T. H., Smirnov, D. S., Salmon, D. P., & Galasko, D. (2020). Failure to stop autocorrect errors in reading aloud increases in aging especially with a positive biomarker for Alzheimer’s disease. *Psychology and Aging*, 35(7), 1016–1025. <https://doi.org/10.1037/pag0000550>
- Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes*, 25(2), 149–188. <https://doi.org/10.1080/01690960902965951>
- Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General*, 121(4), 480–506. <https://doi.org/10.1037/0096-3445.121.4.480>
- Gray, J. R., Burgess, G. C., Schaefer, A., Yarkoni, T., Larsen, R. J., & Braver, T. S. (2005). Affective personality differences in neural processing efficiency confirmed using fMRI. *Cognitive, Affective & Behavioral Neuroscience*, 5(2), 182–190. <https://doi.org/10.3758/CABN.5.2.182>
- Gregg, J., & Inhoff, A. W. (2016). Misperception of orthographic neighbors during silent and oral reading. *Journal of Experimental Psychology: Human Perception and Performance*, 42(6), 799–820. <https://doi.org/10.1037/xhp0000193>
- Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, 6(4), 430–450. <https://doi.org/10.1037/1082-989X.6.4.430>
- Hacker, D. J., Dunlosky, J., & Graesser, A. C. (1998). *Metacognition in educational theory and practice*. Routledge. <https://doi.org/10.4324/9781410602350>
- Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8(4), 439–483. <https://doi.org/10.1080/01690969308407585>
- Haier, R. J., Siegel, B., Tang, C., Abel, L., & Buchsbaum, M. S. (1992). Intelligence and changes in regional cerebral glucose metabolic rate following learning. *Intelligence*, 16(3–4), 415–426. [https://doi.org/10.1016/0160-2896\(92\)90018-M](https://doi.org/10.1016/0160-2896(92)90018-M)
- Hakala, C. M., & O’Brien, E. J. (1995). Strategies for resolving coherence breaks in reading. *Discourse Processes*, 20(2), 167–185. <https://doi.org/10.1080/01638539509544936>

- Helder, A., Van Leijenhorst, L., & van den Broek, P. (2016). Coherence monitoring by good and poor comprehenders in elementary school: Comparing offline and online measures. *Learning and Individual Differences*, 48, 17–23. <https://doi.org/10.1016/j.lindif.2016.02.008>
- Henderson, D., Poppe, A. B., Barch, D. M., Carter, C. S., Gold, J. M., Ragland, J. D., Silverstein, S. M., Strauss, M. E., & MacDonald, A. W., III. (2012). Optimization of a goal maintenance task for use in clinical applications. *Schizophrenia Bulletin*, 38(1), 104–113. <https://doi.org/10.1093/schbul/sbr172>
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179–185. <https://doi.org/10.1007/BF02289447>
- Howe, A. E., Arnell, K. M., Klein, R. M., Joanisse, M. F., & Tannock, R. (2006). The ABCs of computerized naming: Equivalency, reliability, and predictive validity of a computerized rapid automatized naming (RAN) task. *Journal of Neuroscience Methods*, 151(1), 30–37. <https://doi.org/10.1016/j.jneumeth.2005.07.014>
- Hsu, N. S., Jaeggi, S. M., & Novick, J. M. (2017). A common neural hub resolves syntactic and non-syntactic conflict through cooperation with task-specific networks. *Brain and Language*, 166, 63–77. <https://doi.org/10.1016/j.bandl.2016.12.006>
- Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *The American Journal of Psychiatry*, 167(7), 748–751. <https://doi.org/10.1176/appi.ajp.2010.09091379>
- Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience*, 15(1), 98–110. <https://doi.org/10.1162/089892903321107855>
- Kaufman, R. C., & Obler, L. K. (1995). Classification of Normal Reading Error Types. In C. K. Leong & R. M. Joshi (Eds.), *Developmental and acquired dyslexia: Neuropsychological and neurolinguistic perspectives* (pp. 149–157). Springer. https://doi.org/10.1007/978-94-017-1241-5_10
- Kerns, J. G., Cohen, J. D., MacDonald, A. W., III Johnson, M. K., Stenger, V. A., Aizenstein, H., & Carter, C. S. (2005). Decreased conflict- and error-related activity in the anterior cingulate cortex in subjects with schizophrenia. *The American Journal of Psychiatry*, 162(10), 1833–1839. <https://doi.org/10.1176/appi.ajp.162.10.1833>
- 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>
- Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52(2), 205–225. <https://doi.org/10.1016/j.jml.2004.10.002>
- Kinnunen, R., & Vauras, M. (1995). Comprehension monitoring and the level of comprehension in high- and low-achieving primary school children's reading. *Learning and Instruction*, 5(2), 143–165. [https://doi.org/10.1016/0959-4752\(95\)00009-R](https://doi.org/10.1016/0959-4752(95)00009-R)
- Kolk, H., & Chwilla, D. (2007). Late positivities in unusual situations. *Brain and Language*, 100(3), 257–261. <https://doi.org/10.1016/j.bandl.2006.07.006>
- Koolen, S., Vissers, C., Th, W. M., Egger, J. I. M., & Verhoeven, L. (2013). Can monitoring in language comprehension in Autism Spectrum Disorder be modulated? Evidence from event-related potentials. *Biological Psychology*, 94(2), 354–368. <https://doi.org/10.1016/j.biopsycho.2013.07.008>
- Kos, M., van den Brink, D., & Hagoort, P. (2012). Individual variation in the late positive complex to semantic anomalies. *Frontiers in Psychology*, 3, 318. <https://doi.org/10.3389/fpsyg.2012.00318>
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146, 23–49. <https://doi.org/10.1016/j.brainres.2006.12.063>
- Kuperberg, G. R. (2010). Language in schizophrenia Pt. 1: An Introduction. *Language and Linguistics Compass*, 4(8), 576–589. <https://doi.org/10.1111/j.1749-818X.2010.00216.x>
- Kuperberg, G. R., Brothers, T., & Wlotko, E. W. (2020). A Tale of two positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience*, 32(1), 12–35. https://doi.org/10.1162/jocn_a_01465
- Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32–59. <https://doi.org/10.1080/23273798.2015.1102299>
- Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain and Language*, 100(3), 223–237. <https://doi.org/10.1016/j.bandl.2005.12.006>
- Kuperberg, G. R., McGuire, P. K., & David, A. S. (1998). Reduced sensitivity to linguistic context in schizophrenic thought disorder: Evidence from on-line monitoring for words in linguistically anomalous sentences. *Journal of Abnormal Psychology*, 107(3), 423–434. <https://doi.org/10.1037/0021-843X.107.3.423>
- Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, 17(1), 117–129. [https://doi.org/10.1016/S0926-6410\(03\)00086-7](https://doi.org/10.1016/S0926-6410(03)00086-7)
- Kuperberg, G. R., Sitnikova, T., Goff, D., & Holcomb, P. J. (2006). Making sense of sentences in schizophrenia: Electrophysiological evidence for abnormal interactions between semantic and syntactic processing. *Journal of Abnormal Psychology*, 115(2), 251–265. <https://doi.org/10.1037/0021-843X.115.2.251>
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62(1), 621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), 211–240. <https://doi.org/10.1037/0033-295X.104.2.211>
- Larson, G. E., Haier, R. J., LaCasse, L., & Hazen, K. (1995). Evaluation of a “mental effort” hypothesis for correlations between cortical metabolism and intelligence. *Intelligence*, 21(3), 267–278. [https://doi.org/10.1016/0160-2896\(95\)90017-9](https://doi.org/10.1016/0160-2896(95)90017-9)
- Lesh, T. A., Niendam, T. A., Minzenberg, M. J., & Carter, C. S. (2011). Cognitive control deficits in schizophrenia: Mechanisms and meaning. *Neuropsychopharmacology*, 36(1), 316–338. <https://doi.org/10.1038/npp.2010.156>
- Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14(1), 41–104. [https://doi.org/10.1016/0010-0277\(83\)90026-4](https://doi.org/10.1016/0010-0277(83)90026-4)
- Luna, B., Padmanabhan, A., & O'Hearn, K. (2010). What has fMRI told us about the development of cognitive control through adolescence? *Brain and Cognition*, 72(1), 101–113. <https://doi.org/10.1016/j.bandc.2009.08.005>
- Luna, B., Thulborn, K. R., Munoz, D. P., Merriam, E. P., Garver, K. E., Minshew, N. J., Keshavan, M. S., Genovese, C. R., Eddy, W. F., & Sweeney, J. A. (2001). Maturation of widely distributed brain function subserves cognitive development. *NeuroImage*, 13(5), 786–793. <https://doi.org/10.1006/nimg.2000.0743>
- Lyon, G. R. (1995). Toward a definition of dyslexia. *Annals of Dyslexia*, 45(1), 1–27. <https://doi.org/10.1007/BF02648210>
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109(2), 163–203. <https://doi.org/10.1037/0033-2909.109.2.163>

- Mattay, V. S., Fera, F., Tessitore, A., Hariri, A. R., Berman, K. F., Das, S., Meyer-Lindenberg, A., Goldberg, T. E., Callicott, J. H., & Weinberger, D. R. (2006). Neurophysiological correlates of age-related changes in working memory capacity. *Neuroscience Letters*, 392(1-2), 32–37. <https://doi.org/10.1016/j.neulet.2005.09.025>
- McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology*, 24(2), 222–243. <https://doi.org/10.1037/a0017619>
- McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General*, 141(2), 302–320. <https://doi.org/10.1037/a0025250>
- Metzner, P., von der Malsburg, T., Vassilth, S., & Rösler, F. (2017). The importance of reading naturally: Evidence from combined recordings of eye movements and electric brain potentials. *Cognitive Science*, 41(6), 1232–1263. <https://doi.org/10.1111/cogs.12384>
- Manoach, D. S. (2003). Prefrontal cortex dysfunction during working memory performance in schizophrenia: Reconciling discrepant findings. *Schizophrenia Research*, 60(2–3), 285–298. [https://doi.org/10.1016/S0920-9964\(02\)00294-3](https://doi.org/10.1016/S0920-9964(02)00294-3)
- Münter, T. F., Heinze, H.-J., Matzke, M., Wieringa, B. M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia*, 36(3), 217–226. [https://doi.org/10.1016/S0028-3932\(97\)00119-X](https://doi.org/10.1016/S0028-3932(97)00119-X)
- Nakano, H., Saron, C., & Swaab, T. Y. (2010). Speech and span: Working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *Journal of Cognitive Neuroscience*, 22(12), 2886–2898. <https://doi.org/10.1162/jocn.2009.21400>
- Nakatani, C. H., & Hirschberg, J. (1994). A corpus-based study of repair cues in spontaneous speech. *The Journal of the Acoustical Society of America*, 95(3), 1603–1616. <https://doi.org/10.1121/1.408547>
- Ni, W., Fodor, J. D., Crain, S., & Shankweiler, D. (1998). Anomaly detection: Eye movement patterns. *Journal of Psycholinguistic Research*, 27(5), 515–539. <https://doi.org/10.1023/A:1024996828734>
- Niendam, T. A., Laird, A. R., Ray, K. L., Dean, Y. M., Glahn, D. C., & Carter, C. S. (2012). Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cognitive, Affective & Behavioral Neuroscience*, 12(2), 241–268. <https://doi.org/10.3758/s13415-011-0083-5>
- Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin*, 131(4), 510–532. <https://doi.org/10.1037/0033-2909.131.4.510>
- Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology*, 38(5), 752–760. <https://doi.org/10.1111/1469-8986.3850752>
- Nieuwland, M. S., & Van Berkum, J. J. A. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research*, 24(3), 691–701. <https://doi.org/10.1016/j.cogbrainres.2005.04.003>
- Nooteboom, S. G. (1980). Speaking and unspeaking: Detection and correction of phonological and lexical errors in spontaneous speech. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen and hand* (pp. 87–95). Academic Press.
- 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>
- Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology*, 63(1), 1–33. <https://doi.org/10.1016/j.cogpsych.2011.05.001>
- Nozari, N., & Thompson-Schill, S. L. (2016). Chapter 46—Left ventrolateral prefrontal cortex in processing of words and sentences. In G. Hickok & S. L. Small (Eds.), *Neurobiology of language* (pp. 569–584). Academic Press. <https://doi.org/10.1016/B978-0-12-407794-2.00046-8>
- Oakhill, J. V., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading*, 16(2), 91–121. <https://doi.org/10.1080/10888438.2010.529219>
- Oakhill, J., & Yuill, N. (1996). Higher order factors in comprehension disability: Processes and remediation. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 69–92). Lawrence Erlbaum Associates Publishers.
- Oakhill, J., Hartt, J., & Samols, D. (2005). Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and Writing*, 18(7), 657–686. <https://doi.org/10.1007/s11145-005-3355-z>
- Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31(6), 785–806. [https://doi.org/10.1016/0749-596X\(92\)90039-Z](https://doi.org/10.1016/0749-596X(92)90039-Z)
- Osterhout, L., Kim, A., & Kuperberg, G. R. (2012). The neurobiology of sentence comprehension. *The Cambridge handbook of psycholinguistics* (pp. 365–389). Cambridge University Press. <https://doi.org/10.1017/CBO9781139029377.019>
- Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A means for exploring the neurocognition of second language processing. *Language Learning*, 56(s1), 199–230. <https://doi.org/10.1111/j.1467-9922.2006.00361.x>
- Paczynski, M., & Kuperberg, G. R. (2012). Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *Journal of Memory and Language*, 67(4), 426–448. <https://doi.org/10.1016/j.jml.2012.07.003>
- Payne, B. R., Stites, M. C., & Federmeier, K. D. (2019). Event-related brain potentials reveal how multiple aspects of semantic processing unfold across parafoveal and foveal vision during sentence reading. *Psychophysiology*, 56(10), e13432. <https://doi.org/10.1111/psyp.13432>
- Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(6), 1290–1301. <https://doi.org/10.1037/0278-7393.30.6.1290>
- Rubman, C. N., & Salatas Waters, H. (2000). A, B seeing: The role of constructive processes in children's comprehension monitoring. *Journal of Educational Psychology*, 92(3), 503–514. <https://doi.org/10.1037/0022-0663.92.3.503>
- Rypma, B., Berger, J. S., Prabhakaran, V., Bly, B. M., Kimberg, D. Y., Biswal, B. B., & D'Esposito, M. (2006). Neural correlates of cognitive efficiency. *NeuroImage*, 33(3), 969–979. <https://doi.org/10.1016/j.neuroimage.2006.05.065>
- Salthouse, T. A. (1988). The Role of processing resources in cognitive aging. In M. L. Howe & C. J. Brainerd (Eds.), *Cognitive development in adulthood: Progress in cognitive development research* (pp. 185–239). Springer. https://doi.org/10.1007/978-1-4612-3852-2_7
- Sanford, A. J., Leuthold, H., Bohan, J., & Sanford, A. J. S. (2011). Anomalies at the borderline of awareness: An ERP study. *Journal of Cognitive Neuroscience*, 23(3), 514–523. <https://doi.org/10.1162/jocn.2009.21370>
- Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews Neuroscience*, 10(3), 211–223. <https://doi.org/10.1038/nrn2573>
- Sassenhagen, J., & Fiebach, C. J. (2019). Finding the P3 in the P600: Decoding shared neural mechanisms of responses to syntactic violations and oddball targets. *NeuroImage*, 200, 425–436. <https://doi.org/10.1016/j.neuroimage.2019.06.048>

- Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, 137, 29–39. <https://doi.org/10.1016/j.bandl.2014.07.010>
- Schotter, E. R., Tran, R., & Rayner, K. (2014). Don't believe what you read (only once): Comprehension is supported by regressions during reading. *Psychological Science*, 25(6), 1218–1226. <https://doi.org/10.1177/0956797614531148>
- Servan-Schreiber, D., Cohen, J. D., & Steingard, S. (1996). Schizophrenic deficits in the processing of context. A test of a theoretical model. *Archives of General Psychiatry*, 53(12), 1105–1112. <https://doi.org/10.1001/archpsyc.1996.01830120037008>
- Simonsohn, U. (2018). Two lines: A valid alternative to the invalid testing of u-shaped relationships with quadratic regressions. *Advances in Methods and Practices in Psychological Science*, 1(4), 538–555. <https://doi.org/10.1177/2515245918805755>
- Singer, J. K., Lichtenberger, E. O., Kaufman, J. C., Kaufman, A. S., & Kaufman, N. L. (2012). The Kaufman assessment battery for children—2nd ed. and the Kaufman test of educational achievement—2nd ed. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 269–296). Guilford Press.
- Solomon, M., Ozonoff, S. J., Cummings, N., & Carter, C. S. (2008). Cognitive control in autism spectrum disorders. *International Journal of Developmental Neuroscience*, 26(2), 239–247. <https://doi.org/10.1016/j.ijdevneu.2007.11.001>
- South, M., Larson, M. J., Krauskopf, E., & Clawson, A. (2010). Error processing in high-functioning. *Autism Spectrum Disorders. Biological Psychology*, 85(2), 242–251. <https://doi.org/10.1016/j.biopsycho.2010.07.009>
- Stanovich, K. E., & Cunningham, A. E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory & Cognition*, 20(1), 51–68. <https://doi.org/10.3758/BF03208254>
- Suggate, S. P. (2016). A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities*, 49(1), 77–96. <https://doi.org/10.1177/0022219414528540>
- Szewczyk, J. M., & Schriefers, H. (2011). Is animacy special? ERP correlates of semantic violations and animacy violations in sentence processing. *Brain Research*, 1368, 208–221. <https://doi.org/10.1016/j.brainres.2010.10.070>
- Tanner, D. (2019). Robust neurocognitive individual differences in grammatical agreement processing: A latent variable approach. *Cortex*, 111, 210–237. <https://doi.org/10.1016/j.cortex.2018.10.011>
- Tanner, D., Goldshtein, M., & Weissman, B. (2018). Individual differences in the real-time neural dynamics of language comprehension. In K. D. Federmeier & D. G. Watson (Eds.), *Psychology of learning and motivation* (Vol. 68, pp. 299–335). Academic Press. <https://doi.org/10.1016/bs.plm.2018.08.007>
- Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence*. *Bilingualism: Language and Cognition*, 16(2), 367–382. <https://doi.org/10.1017/S1366728912000302>
- Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia*, 56, 289–301. <https://doi.org/10.1016/j.neuropsychologia.2014.02.002>
- Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *The Journalism Quarterly*, 30(4), 415–433. <https://doi.org/10.1177/107769905303000401>
- Torgesen, J. K., Wagner, R., & Rashotte, C. (2012). *Test of Word Reading Efficiency: (TOWRE-2)*. Pearson Clinical Assessment.
- Traxler, M. J. (2014). Trends in syntactic parsing: Anticipation, Bayesian estimation, and good-enough parsing. *Trends in Cognitive Sciences*, 18(11), 605–611. <https://doi.org/10.1016/j.tics.2014.08.001>
- Traxler, M. J., Foss, D. J., Seely, R. E., Kaup, B., & Morris, R. K. (2000). Priming in sentence processing: Intralexical spreading activation, schemas, and situation models. *Journal of Psycholinguistic Research*, 29(6), 581–595. <https://doi.org/10.1023/A:1026416225168>
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37(3), 498–505. <https://doi.org/10.3758/BF03192720>
- Uttl, B. (2002). North American Adult Reading Test: Age norms, reliability, and validity. *Journal of Clinical and Experimental Neuropsychology*, 24(8), 1123–1137. <https://doi.org/10.1076/jcen.24.8.1123.8375>
- van de Meerendonk, N. V. D., Kolk, H. H. J., Chwilla, D. J., & Vissers, C. T. W. M. (2009). Monitoring in language perception. *Language and Linguistics Compass*, 3(5), 1211–1224. <https://doi.org/10.1111/j.1749-818X.2009.00163.x>
- van de Meerendonk, N., Kolk, H. H., Vissers, C. T., & Chwilla, D. J. (2010). Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, 22(1), 67–82. <https://doi.org/10.1162/jocn.2008.21170>
- van de Meerendonk, N., Rueschemeyer, S.-A., & Kolk, H. H. J. (2013). Language comprehension interrupted: Both language errors and word degradation activate Broca's area. *Brain and Language*, 126(3), 291–301. <https://doi.org/10.1016/j.bandl.2013.07.004>
- van Veen, V., & Carter, C. S. (2002). The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiology & Behavior*, 77(4–5), 477–482. [https://doi.org/10.1016/S0031-9384\(02\)00930-7](https://doi.org/10.1016/S0031-9384(02)00930-7)
- Vaughn, S., Levy, S., Coleman, M., & Bos, C. S. (2002). Reading instruction for students with LD and EBD: A synthesis of observation studies. *The Journal of Special Education*, 36(1), 2–13. <https://doi.org/10.1177/00224669020360010101>
- Vazey, E. M., Moorman, D. E., & Aston-Jones, G. (2018). Phasic locus coeruleus activity regulates cortical encoding of salience information. *Proceedings of the National Academy of Sciences of the United States of America*, 115(40), E9439–E9448. <https://doi.org/10.1073/pnas.1803716115>
- Vissers, C. T. W. M., Chwilla, D. J., & Kolk, H. H. J. (2006). Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. *Brain Research*, 1106(1), 150–163. <https://doi.org/10.1016/j.brainres.2006.05.012>
- Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, 428(6984), 748–751. <https://doi.org/10.1038/nature02447>
- 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>
- Wagoner, S. A. (1983). Comprehension monitoring: What it is and what we know about it. *Reading Research Quarterly*, 18(3), 328–346. <https://doi.org/10.2307/747392>
- Wang, L., Schoot, L., Brothers, T., Alexander, E., Warnke, L., Kim, M., Khan, S., Hämmäläinen, M., & Kuperberg, G. R. (2021). *Dynamic predictive coding across the left fronto-temporal language hierarchy: Evidence from MEG, EEG and fMRI* (p. 2021.02.17.431452). <https://doi.org/10.1101/2021.02.17.431452>
- Warren, T., & McConnell, K. (2007). Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic Bulletin & Review*, 14(4), 770–775. <https://doi.org/10.3758/BF03196835>
- Warren, T., Milburn, E., Patson, N. D., & Dickey, M. W. (2015). Comprehending the impossible: What role do selectional restriction violations play? *Language, Cognition and Neuroscience*, 30(8), 932–939. <https://doi.org/10.1080/23273798.2015.1047458>

- Wolf, M. (1986). Rapid alternating stimulus naming in the developmental dyslexias. *Brain and Language*, 27(2), 360–379. [https://doi.org/10.1016/0093-934X\(86\)90025-8](https://doi.org/10.1016/0093-934X(86)90025-8)
- Woodcock, R. W. (1973). *Woodcock reading mastery tests*. American Guidance Service.
- Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock–Johnson Revised tests of cognitive ability: Standard and supplemental batteries*. Riverside Publishing.
- Yarkoni, T., & Braver, T. S. (2010). Cognitive neuroscience approaches to individual differences in working memory and executive control: Conceptual and methodological issues. In A. Gruszka, G. Matthews, & B. Szymura (Eds.), *Handbook of individual differences in cognition: Attention, memory, and executive control* (pp. 87–107). Springer. https://doi.org/10.1007/978-1-4419-1210-7_6
- Ye, Z., & Zhou, X. (2008). Involvement of cognitive control in sentence comprehension: Evidence from ERPs. *Brain Research*, 1203, 103–115. <https://doi.org/10.1016/j.brainres.2008.01.090>
- Ye, Z., & Zhou, X. (2009). Conflict control during sentence comprehension: fMRI evidence. *NeuroImage*, 48(1), 280–290. <https://doi.org/10.1016/j.neuroimage.2009.06.032>
- Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, 111(4), 931–959. <https://doi.org/10.1037/0033-295x.111.4.939>
- Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, 46(4), 681–692. <https://doi.org/10.1016/j.neuron.2005.04.026>
- Yu, A. J., Dayan, P., & Cohen, J. D. (2009). Dynamics of attentional selection under conflict: Toward a rational Bayesian account. *Journal of Experimental Psychology: Human Perception and Performance*, 35(3), 700–717. <https://doi.org/10.1037/a0013553>
- Yudes, C., Macizo, P., Morales, L., & Bajo, M. T. (2013). Comprehension and error monitoring in simultaneous interpreters. *Applied Psycholinguistics*, 34(5), 1039–1057. <https://doi.org/10.1017/S0142716412000112>
- Zabracky, K., & Ratner, H. H. (1989). Effects of Reading Ability on Children's Comprehension Evaluation and Regulation. *Journal of Reading Behavior*, 21(1), 69–83. <https://doi.org/10.1080/10862968909547659>
- Zheng, X., & Lemhöfer, K. (2019). The “semantic P600” in second language processing: When syntax conflicts with semantics. *Neuropsychologia*, 127, 131–147. <https://doi.org/10.1016/j.neuropsychologia.2019.02.010>

Received November 6, 2020

Revision received July 15, 2021

Accepted July 22, 2021 ■