Selective effects of cognitive control on the P600: Evidence from a large-scale individual differences study

Trevor Brothers, Eddie Wlotko, Margarita Zeitlin, Arim Choi Perrachione, Connie Choi, Minjae Kim, Gina Kuperberg
Detecting Language Errors

• Linguistic errors are an everyday part of comprehension *Fromkin, 1971*

• Errors can arise from the producer, or the comprehender due to misinterpretations or perceptual errors
  • E.g. misperceived orthographic neighbors

• Successful comprehension requires
  1. rapidly *detecting* conflict between the bottom-up input and contextual constraints
  2. engaging in behaviors to help resolve this conflict (e.g. regressive saccades)

I would like to spank all teachers.
-George W. Bush

The President is going to lead us out of this recovery.
-Dan Quayle
The P600 is elicited following syntactic errors or syntactically dis-preferred continuations. 

Osterhout & Holcomb, 1992

The P600 is also elicited to semantic anomalies, even when no compelling syntactic “edit” is available. 


Robust P600s are only elicited when anomalies are consciously detected. 

Sanford et al., 2010; Batterink & Neville, 2013
Individual Differences in Error Detection?

- Error detection ability increases during development throughout childhood

- Error detection in self-generated speech and comprehension is impaired in schizophrenia Kuperberg et al., 1998

- Impaired error detection during discourse comprehension in multiple developmental disorders (ADHD, SLI)

Walczyk & Raska, 1992

Hanley et al. 2016

McInnes, Humphries, Hogg-Johnson & Tannock, 2003

Englehardt et al., 2010
Individual Differences and the P600

• P600 responses are reduced or absent in lower proficiency L2 learners

• In native speakers, some studies have shown a positive relationship between the semantic P600 and working memory capacity (WMC)

• This relationship is not always present, but a meta-analysis of 4 prior studies shows a small but significant weighted effect

• Less is known about the relationship between working memory capacity and behavioral error detection performance in healthy, native readers

Meta-analysis:
WMC ~ semantic P600s

<table>
<thead>
<tr>
<th>Study</th>
<th>Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nakano, Saron &amp; Swaab, 2010</td>
<td>0.38</td>
</tr>
<tr>
<td>Kim, Oines &amp; Miyake, 2018</td>
<td>0.40</td>
</tr>
<tr>
<td>Kos, van den Brink &amp; Hagoort, 2012</td>
<td>0.12</td>
</tr>
<tr>
<td>Zheng &amp; Lemhofer, 2019</td>
<td>0.23</td>
</tr>
</tbody>
</table>

Weighted ES
Complex span tasks are .... complex

- Complex working memory tasks index multiple cognitive constructs

1) **Memory Processes** – grouping, chunking, rehearsal
2) **Language Experience** – e.g. reading fluency, decoding
3) **Executive Attention/Cognitive Control** – inhibiting distractions, avoiding proactive interference, resolving competition

- Using a Factor Analytic approach, we hope to disentangle the influences of **memory processes**, **language experience**, and **cognitive control** in predicting individual differences in error detection abilities

---

**Complex Span predicts performance in “cognitive control” tasks**

![Bar chart showing Low WM vs High WM for Prosaccade and Antisaccade tasks.](chart)

**Engle, 2002**

**Factor Analysis**

1. Minimize measurement error
2. Minimize task specific variance
3. Dissociate underlying constructs
U-shaped Brain Responses and Cognitive Control

• When examining neural responses, it is important to examine both linear and non-linear relationships.

• BOLD responses in DLPFC show a u-shaped function according to working memory demands as performance breaks down Callicot et al., 1999; Mattay, et al; 2006.

• BOLD responses during the anti-saccade task show u-shaped developmental time-course Luna et al., 2001.
  - Childhood – low BOLD, low acc.
  - Adolescence – high BOLD, med acc.
  - Adulthood – low BOLD, high acc.
Analysis Plan

ERP Discourse Comprehension Task + Individual Differences Battery

Filtering, Artifact Rejection, Averaging

Extract *mean ERP responses* in each spatio-temporal ROI

Outlier removal, and Factor Analysis

Extract *Factor Scores* for each latent construct of interest

Use factor scores to predict mean ERP difference scores via multiple regression
Processing semantic errors in ERP/MEG

- Participants read 3-sentence discourse scenarios while EEG was recorded
  - Performed delayed plausibility judgment for each discourse
- Critical nouns were either congruent or semantically anomalous
  - 50% animate, 50% inanimate, counterbalanced to produce animacy violations
- Final sentence presented RSVP (550ms per word)
- P600 amplitudes were measured over central-posterior electrode sites (600-1000ms)

### ERP Anomaly Effect (N=77)

The lifeguards heard a report of a shark right near the beach.

Their first priority was to prevent an incidents in the sea.

Hence, they cautioned the trainees/drawer…

- **N400**
  - 300-500ms
- **P600**
  - 600-1000ms

---

**Congruent**

**Anomalous**
Individual Differences Battery

- **Working memory Capacity**
  - Reading Span
  - Listening Span
  - Operation Span
  - Subtract-2 Span

- **Language Experience**
  - North American Adult Reading Test
  - WRMT – Word ID
  - Author recognition

- **Cognitive Control**
  - AX- Continuous Performance Task
    - AY Accuracy
    - AY – AX RT cost
  - Manual Stroop Task
    - Incong Acc.

  ![Diagram of AX-Continuous Performance Task](image)

- Internal consistency metrics were good to acceptable (0.89 – 0.62; mean $\alpha = 0.82$)
- Within-construct correlations were robust (mean $r = 0.51$) and exceeded between-construct correlations (mean $r = 0.26$)
Exploratory Factor Analysis

- All measures included in an Exploratory Factor Analysis
  - Maximum Likelihood
  - Varimax Rotation (orthogonal predictors)
- Resulted a three-factor solution, with relatively low cross-loadings between factors
- Demonstrates convergent and divergent validity across domains
- Regression-based factors scores were normally distributed
- Linear and non-linear relationships were examined using multiple regression

<table>
<thead>
<tr>
<th></th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSPAN</td>
<td>0.75</td>
<td>0.22</td>
<td>0.05</td>
</tr>
<tr>
<td>Subt-2 SPAN</td>
<td>0.69</td>
<td>0.11</td>
<td>0.21</td>
</tr>
<tr>
<td>RSPAN</td>
<td>0.66</td>
<td>0.15</td>
<td>0.01</td>
</tr>
<tr>
<td>LSPAN</td>
<td>0.63</td>
<td>0.24</td>
<td>0.07</td>
</tr>
<tr>
<td>ART</td>
<td>0.12</td>
<td>0.64</td>
<td>0.05</td>
</tr>
<tr>
<td>NAART</td>
<td>0.37</td>
<td>0.89</td>
<td>0.16</td>
</tr>
<tr>
<td>WORD_ID</td>
<td>0.43</td>
<td>0.58</td>
<td>0.37</td>
</tr>
<tr>
<td>AY accuracy</td>
<td>0.06</td>
<td>0.13</td>
<td>0.81</td>
</tr>
<tr>
<td>AY-AX RT</td>
<td>0.06</td>
<td>0.05</td>
<td>0.70</td>
</tr>
<tr>
<td>Stroop Acc.</td>
<td>0.27</td>
<td>0.13</td>
<td>0.28</td>
</tr>
</tbody>
</table>
Predicting Behavioral Accuracy ($d'$)

Multiple Regression Model, $F(6,76) = 3.0$, $R^2 = .21^*$

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>SE</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>2.8</td>
<td>0.11</td>
<td>25.01</td>
<td>&gt; 0.001</td>
</tr>
<tr>
<td>WMC</td>
<td>0.07</td>
<td>0.08</td>
<td>0.93</td>
<td>0.35</td>
</tr>
<tr>
<td>Lang. Exp.</td>
<td>0.06</td>
<td>0.08</td>
<td>0.78</td>
<td>0.44</td>
</tr>
<tr>
<td>Control</td>
<td><strong>0.18</strong></td>
<td><strong>0.08</strong></td>
<td><strong>2.20</strong></td>
<td><strong>0.03</strong> *</td>
</tr>
<tr>
<td>WMC$^2$</td>
<td>-0.09</td>
<td>0.07</td>
<td>-1.30</td>
<td>0.20</td>
</tr>
<tr>
<td>Lang. Exp.$^2$</td>
<td>-0.10</td>
<td>0.05</td>
<td>-1.87</td>
<td>0.07</td>
</tr>
<tr>
<td>Control$^2$</td>
<td>-0.05</td>
<td>0.07</td>
<td>-0.69</td>
<td>0.49</td>
</tr>
</tbody>
</table>

For the linear model:
- $b = 0.18^*$

For the quadratic model:
- $r = .15$
- $r = .19$
- $r = .29^{**}$

Graphs showing the relationship between factors and $d'$ with appropriate correlation coefficients.
Predicting P600 Amplitudes

Multiple Regression Model, $F(6,76) = 2.5, R^2 = .18^*$

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>SE</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.39</td>
<td>0.58</td>
<td>5.86</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>WMC</td>
<td>-0.46</td>
<td>0.41</td>
<td>0.26</td>
<td>0.26</td>
</tr>
<tr>
<td>Lang. Exp.</td>
<td>-0.01</td>
<td>0.39</td>
<td>-0.04</td>
<td>0.97</td>
</tr>
<tr>
<td>Control</td>
<td>0.12</td>
<td>0.43</td>
<td>0.28</td>
<td>0.78</td>
</tr>
<tr>
<td>WMC$^2$</td>
<td>-0.17</td>
<td>0.36</td>
<td>-0.45</td>
<td>0.65</td>
</tr>
<tr>
<td>Lang. Exp.$^2$</td>
<td>-0.02</td>
<td>0.28</td>
<td>-0.05</td>
<td>0.96</td>
</tr>
<tr>
<td>Control$^2$</td>
<td>-1.31</td>
<td>0.37</td>
<td>-3.52</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

$r = -.06$

$r = .02$

$r = .40^{***}$
Examining the U-shaped relationship

- After dividing participants, readers with moderate levels of cognitive control showed larger P600 effects than either high or low control participants
  - Driven by ERP differences for anomalous words

- An inverse relationship was observed for the N400
  - Potential “tradeoff relationship” between the N400/P600
  - May also reflect overlap of opposite polarity ERP components
The amplitude of the P600 appears to depend on both the *effectiveness* and *efficiency* of cognitive control mechanisms.

- **Effectiveness** – Low control participants appear to show impairments in *conflict detection*, with undetected semantic anomalies resulting in a reduced P600.

- **Efficiency** – Readers with very high control abilities can *detect* and *resolve* language conflicts more efficiently, resulting in a smaller P600 responses.

Explaining the U-shaped relationship
Take-home messages

• Domain general, cognitive control abilities are an important source of variability in comprehension performance among native adult readers.

• The relationship between neural responses and underlying cognitive constructs is not always a simple linear function.

• Factor Analytic techniques (and SEM) are powerful tools
  • Measure multiple constructs using multiple tasks!

• Complex span tasks are multifaceted – they are often good predictors, but poor for identifying underlying cognitive mechanisms.
THANK YOU!

• Thanks to NSF and everyone in the Neurocognition of Language Lab for their support!

• Special thanks to: Arjun Dutta, Emma Fleisher, Sophie Greene, Sayaka Koga, Nicole Nadwodny, Nimarta Narang, Rebecca Nardulli, Chelsey Ott, Ola Ozernov-Palchik, Simone Riley, Rebecca Schaub, and Lena Warnke!

Trevor Brothers, Eddie Wlotko, Margarita Zeitlin, Arim Choi Perrachione, Connie Choi, Minjae Kim, Gina Kuperberg
The basic claim of the trade-off hypothesis is that readers can engage in *either* attempted semantic access (N400) *or* structural re-analysis (P600) following an anomaly

- “The hearty meal was *devouring*…”
- Semantic access is not optional, and animacy violations have no obvious structural repair
  - “They cautioned the *drawer*…”
- Reduced N400s for readers with large P600s can also be explained via component overlap
- Results from other neuroimaging methods (MEG, fMRI) will be necessary to resolve this issue
This relationship ($r = 0.40, p < 0.001$), remained highly significant after Bonferroni correction for multiple comparisons.

And remained significant after controlling for an additional measures related to “processing speed.”

Two-Line Test – this relationship was truly U-shaped, not just quadratic (Simonsohn, 2018; http://webstimate.org/twolines/)
Thoughts on Measuring Cognitive Control

- Differences in cognitive control are notoriously difficult to measure (Paap & Sawi, 2014; Hedge, et al., 2018)
  - Low internal reliability
  - Low inter-task correlations
  - Biased from global RT differences

- Measurements based on difference scores produce high measurement error (and may eliminate effects of interest)

- Accuracy measures are often at ceiling, with little reliable variation in healthy adults

- Make sure tasks are sufficiently difficult, with large numbers of trials (100+)

Better tasks
- Stroop
- AX-CPT (short ISI)
- Anti-saccade

Worse tasks
- Flanker
- Stop Signal
- Switching Tasks

Worst tasks
- Negative priming
- Simon
Individual Differences: Language Experience

- **Language Experience/Vocabulary**
  - North American Adult Reading Test
  - WRMT – Word ID
  - Author recognition

- Two reading comprehension tasks were also administered (*KTEA, WRMT passages*)

- These loaded equally on measures of *language experience* (*r = .49*) and *working memory* (*r = 0.46*)

- Therefore, they were not included in the Factor Analysis

<table>
<thead>
<tr>
<th>Vocabulary</th>
<th>Author Recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOUQUET</td>
<td>F. Scott Fitzgerald</td>
</tr>
<tr>
<td>DEPOT</td>
<td>Kristen Steinke</td>
</tr>
<tr>
<td>GIST</td>
<td>Harper Lee</td>
</tr>
<tr>
<td>LEVIATHAN</td>
<td>Isaac Asimov</td>
</tr>
<tr>
<td>GAUCHE</td>
<td>S. L. Halloway</td>
</tr>
<tr>
<td>BEATIFY</td>
<td>Jessica Ann Lewis</td>
</tr>
</tbody>
</table>

Inter-correlations ranged from *r = .78 - .43*
Individual Differences: Working Memory Capacity

- **Linguistic WMC**
  - Reading Span
  - Listening Span
- **Non-Linguistic WMC**
  - Operation Span
  - Subtract-2 Span

- All of these tasks were “complex span” measures, that interleaved processing and storage demands

- Have been shown to better predict complex behaviors and fluid intelligence

Inter-correlations ranged from \( r = .65 - .49 \)
Individual Differences: Cognitive Control

**AX-Continuous Performance Task**
- AY Trial Accuracy
- AY vs. AX reaction time cost

**Manual Stroop Task**
- Incongruent Trial Accuracy
  - congruent: 99%, incongruent: 94%
- Incongruent vs. Cong RT cost

**AX-CPT Overall Results**
- Inter-correlations ranged from $r = .56 - .23$