Conflict Adaptation between Stroop and Sentence Comprehension: An Eye Tracking Study

by Christine T. Asaro

B.A. in Cognitive Science and Mathematics, May 2013, The Johns Hopkins University

A Thesis submitted to

The Faculty of
The Columbian College of Arts and Sciences
of The George Washington University
in partial fulfillment of the requirements
for the degree of Master of Arts

May 15, 2016

Thesis directed by

Malathi Thothatiri
Assistant Professor of Speech and Hearing

© Copyright 2016 by Christine T. Asaro
All rights reserved

Acknowledgments

The author wishes to express her gratitude to Dr. Malathi Thothathiri, members of the Neuroscience of Language Lab at The George Washington University, and Nina S. Hsu and Jared M. Novick of the University of Maryland for their contributions to this project.

Abstract of Thesis

Conflict Adaptation between Stroop and Sentence Comprehension:
An Eye Tracking Study

Domain-general accounts of executive function posit that cognitive control mechanisms can extend across modules of the brain. Cognitive control mechanisms are relevant to the resolution of conflict, which is the temporary ambiguity caused by different cues giving rise to competing representations. Such temporary misrepresentations exist during Stroop tasks and during some sentence comprehension tasks. This thesis demonstrates a causal connection between cognitive control and syntactic-semantic conflict sentences. Specifically, we found an effect of previous Stroop trial type on sentences containing conflict. On-line eye movement data during sentence comprehension of syntactic-semantic conflict sentences were influenced by control processes activated during the previous Stroop trial only when the Stroop trial also contained conflict. During comprehension of these sentences, looks to the semantically plausible but incorrect image were suppressed. This overlap in conflict processing and control mechanisms between tasks is known as conflict adaptation. We also found a significant correlation between baseline Stroop performance and the amount of adaptation that participants exhibited. Preliminary results indicate that individual differences in performance may be the result of the extent to which a participant relies on semantic cues. Our results show that the activation of cognitive control mechanisms during a previous trial can aid sentence comprehension on a subsequent trial, supporting domain-general accounts of executive functioning.

Table of Contents

Acknowledgments.....iii

Abstract of Thesis.....iv

List of Figures.....vi

List of Tables.....vii

Introduction.....1

Methods.....10

Results.....17

Discussion.....26

References.....32

Appendix.....36

List of Figures

Figure 1.....12

Figure 2.....19

Figure 3.....21

Figure 4.....22

Figure 5.....25

List of Tables

Table 1.....20

Table 2.....20

Table 3.....24

Introduction

There are many theoretical frameworks and models of executive function as a cognitive construct. Among them, functional abilities such as planning, initiation, inhibition, sustained attention, cognitive flexibility, and self-monitoring are often referenced as key components (Constantinidou, Wertheimer, Tsanadis, Evans, & Paul, 20l2). Inhibition, flexibility, and self-monitoring are important whenever a person needs to suppress automatic, routine response patterns in favor of a novel response. For example, consider crossing the street in England – an American pedestrian there will have to consciously inhibit the automatic habit of checking left first for traffic, and look to the right instead. This new behavior is guided by the current context in a top-down way (Miller & Cohen, 2001). Executive function abilities are involved in this and other types of problem solving that involve cognitive control.

The above example is one instance of conflict. Conflict arises in tasks where multiple cues create interference during or between stimulus processing and responding. For example, in the well-known Stroop task, a participant must attend to the ink color of a color word. The representations obtained by identifying the color and by reading the word itself, which also represents a color, may or may not be the same. Conflict occurs when these two representations differ and both become active at the same time. Further processing is necessary to select the correct representation from the two competitors. This is known as representational conflict.

Besides competition at the level of stimulus representation, another instance in which conflict arises is when there is interference created at the response level, which is known as response conflict. In the Stroop example, a participant may need to indicate a

response by pressing a key. The participant has to inhibit the action of pressing the key that represents the interfering response (the color obtained by reading the stimulus word). Thus, the participant has to overcome both conflict created by the differing representations during stimulus processing, as well as conflict at the level of making a response (Kornblum, Hasbroucq, & Osman, 1990).

Conflict also exists in sentence interpretation. For sentences said to contain syntactic-semantic conflict, there is interference between (1) what the reader anticipates to be the meaning of the sentence based on semantic knowledge, and (2) the meaning assigned to the sentence based on its syntactic construction. For example, a sentence such as "the cat was chased by the mouse" contains conflict. There is interference between (1) what a reader knows semantically about cats and mice (and who is probabilistically going to chase whom) and (2) which animal the syntax of the sentence assigns to the agent and patient roles. Thus, conflict is a phenomenon relevant to the linguistic domain and to sentence processing tasks, and is also relevant to conversation and other situations we experience on a daily basis.

To explain how we succeed in tasks that involve conflict, in the lab and in daily life, it has been theorized that monitoring and control mechanisms create biases that support goal-directed behavior (Botvinick, Braver, Barch, Carter, & Cohen, 2001). Botvinick et al. postulate a conflict-control loop in which conflict is first detected by a monitoring system that informs a control system, which then resolves the conflict. Two mechanisms have been proposed to explain conflict resolution: task-relevant information could be selected for increased focus of attention and/or task-irrelevant information could

be suppressed from attention. Each of these mechanisms could aid in overcoming habitual behavior for the resolution of conflict and successful completion of a task.

The frontal lobe has traditionally been implicated as the brain area involved in executive functioning, where conflict monitoring and control mechanisms are located. However, it is not just these frontal brain regions, but also their connections, that are relevant to cognition and behavior. Three fronto-subcortical circuits are implicated in the proper functioning of executive skills. These include the dorso-lateral prefrontal circuit, the orbito-frontal circuit, and the anterior cingulate circuit (Constantinidou et al., 2012). Studies show the anterior cingulate cortex (ACC) to be active during processing of tasks that involve conflict and to predict activity in the prefrontal cortex. Specifically, in a Stroop task, greater activity in the ACC on trials containing conflict was associated with increased control of behavior in the subsequent trial. Additionally, trials following conflict that showed such change in behavior were associated with increased activation of the prefrontal cortex. This indicates that the ACC has a role in monitoring for conflict as well as triggering control mechanisms (Kerns et al., 2004). These areas may have dynamic, complementary roles in conflict resolution and may work as components of a feedback loop (MacDonald, Cohen, Stenger, & Carter, 2000).

Imaging studies provide correlational evidence that frontal regions are involved in conflict control during both sentence processing and non-sentence-related tasks. For example, one study showed that the dorsal medial superior frontal gyrus, left inferior frontal gyrus (LIFG), and left inferior parietal lobe showed greater activation during comprehension of implausible rather than plausible sentences (e.g. “The thief kept the policeman in the station”) and also incongruent rather than congruent Stroop trials (Ye &

Zhou, 2009). Another showed co-localization of activity in the LIFG for Stroop tasks containing representational conflict and for comprehension of sentences containing ambiguity (e.g. "Clean the pig with the leaf") in individual subjects (January, Trueswell, & Thompson-Schill, 2009). A third study explored brain potentials of individuals in low- and in high-control groups during a sentence processing task (Ye & Zhou, 2008). The low-control group demonstrated sustained positivity throughout reanalysis of implausible sentences, suggesting that control mechanisms have sustained activity throughout resolution of conflict in sentence processing and interpretation. Together, these studies provide evidence that Broca's area is not dedicated solely to syntactic processing, because it also has a role in resolving linguistic conflict. This area's activation during conflict resolution in sentence processing as well as non-sentence tasks suggests that the cognitive control mechanisms of executive function and sentence processing do not take place in isolated regions of the brain.

Interestingly, while such control process might be postulated a priori to tax mental resources such that subsequent trials containing conflict would take longer processing time, the opposite effect has been shown. This effect—known as conflict adaptation—involves participants showing *faster* response times on a conflict trial when that trial is preceded by another conflict trial. The conflict adaptation effect is seen in studies that manipulate trial order for tasks that contain congruent and incongruent trials. An incongruent trial is one for which interference exists at either the representational or the response level. A congruent trial does not have such interference. The typical conflict adaptation pattern is that reaction time for an incongruent trial that follows another incongruent trial is faster than that for an incongruent trial that follows a congruent trial.

According to the model by Botvinick and colleagues, this is explained as follows: once conflict is detected in an incongruent trial, control mechanisms are activated and the ability to resolve interference is enhanced on a subsequent incongruent trial (Botvinick et al., 2001). As explained before, this could happen via the facilitation of relevant information and/or the suppression of irrelevant information. At least one study has found support for the latter suppression hypothesis, which is relevant to the interpretation of this study (to come later) (Egner & Hirsch, 2005).

How far does adaptation extend? A domain-general cognitive control system should extend to a variety of domains, including language. Thus, this account would predict that adaptation extends between tasks in different domains because domains of knowledge are interdependent. In contrast, a domain-specific account would predict that the brain is composed of independent modules, such that learning or control mechanisms in one module would not impact another module. If this is the case, adaptation would not be possible between tasks in different domains. The issue is controversial. Studies claiming to support the domain-general view have been attacked on three different counts. First, many studies reporting adaptation effects repeat a small set of stimulus items, which might have given rise to repetition priming rather than adaptation (Mayr, Awh, & Laurey, 2003). Second, even when repetition is controlled (e.g., by switching between multiple stimuli and tasks), Egner argues that executive function cannot be posited to be domain-general *unless the experimental tasks definitively tap different domains and employ different stimuli and response characteristics* (Egner, 2008). Finally, some studies have shown a "switch cost" for switching between domains. If adaptation were generalizable, then task switching after an incongruent trial should be faster than

switching after a congruent trial, no matter what the domains are (Hyafil, Summerfield, & Koechlin, 2009). In sum, the claim that conflict adaptation generalizes between vastly different domains still awaits additional empirical validation.

Some studies have attempted to address the concerns described above. For example, Ullsperger and colleagues only analyzed trials that did not repeat target, distractor, or response and still found an adaptation effect – reaction time and error rates were lower for incompatible trials that immediately followed another incompatible trial (Ullsperger, Bylsma, & Botvinick, 2005). Another study used two tasks: a flanker task that involved response to the direction of arrows and Stroop tasks that did not. The authors found effects of the congruency of the previous task on the current task. Their results cannot be attributed to priming because the entire task changed between trials (Freitas, Bahar, Yang, & Banai, 2007. But see Freitas & Clarke, 2015 for some qualifications). Together, there is some evidence for conflict adaptation across domains, which supports a domain-general view of executive function.

In this thesis, we plan to use the conflict adaptation effect to investigate whether there are causal relationships between domain-general cognitive control and sentence processing. If sentence processing necessarily involves domain-general cognitive control, we would expect conflict adaptation between the Stroop task (which does not involve sentence processing) and sentence comprehension. Previous research has already shown conflict adaptation from sentence processing to the Stroop task (Kan et al., 2013). That is, encountering and resolving conflict during sentence processing facilitates performance on a subsequent Stroop trial. Here, we aim to test for conflict adaptation in the opposite direction to see if encountering and resolving conflict during Stroop trials influences

subsequent sentence processing. This reversal of direction is important because it is more clinically relevant to populations with language disorders.

Typically, conflict adaptation can be seen through differences in reaction time between incongruent trials that follow an incongruent trial as compared to those that follow a congruent trial. Reaction time measures are a noisy and coarse measure of sentence processing. Therefore, we plan to use eye movement measures, which are known to be sensitive to real-time or online processing of sentences. Eye-movement measures have a long history in psycholinguistics in studies of written language and sentence reading. In reading studies, a participant's eyes are tracked as he/she reads and re-reads certain sections of text as needed, in order to fully comprehend a sentence. Then, the time sequence for the correct comprehension of the complex sentence can be traced. Eye movement measures also have been used in the study of spoken language comprehension. For experiments involving auditory presentation of sentences and a visual display that is acted upon, eye tracking provides information about the sequence involved in sentence comprehension beyond simple reaction time, as processing can be examined from one moment to the next in time with incoming information (Rayner, 1998; Tanenhaus & Trueswell, 2006). To summarize, eye movements have been shown to be reliable measures of online sentence processing for sentence reading as well as auditory sentence processing. Further, it has been shown that similar eye movement results are obtained independent of the modality in which people encounter sentences (visual reading versus auditory hearing) (Novick, Thompson-Schill, &Trueswell, 2008).

A preponderance of evidence shows that sentence processing takes multiple sources of information—syntactic, semantic, and even visual contextual factors—into

account. For example, Altmann & Kamide reported that when participants heard a sentence such as "*The man has drunk all of the wine*", their eyes moved to an empty wine glass versus a full beer glass at the point of the verb (Altmann & Kamide, 2007). In other words, listeners do not wait until they hear an object being mentioned. Rather, they use the auditory input and the visual input to *predict* likely targets. This kind of evidence shows that sentence processing is *incremental* (See also Malaia & Newman, 2015).

In the current study, the critical sentences (e.g., *The cat was chased by the mouse*) contained syntactic-semantic conflict. Because sentence processing is incremental, we would predict that as people hear "*The cat was chas…"*, they might reasonably expect that the cat was chasing the mouse (due to semantic plausibility). This would lead them to fixate on the incorrect picture. Upon hearing more of the sentence (e.g., "*The cat was chased by…"*), the listeners will have to revise the original interpretation and look to the correct picture instead. The specific prediction for conflict adaptation is that for sentences like the above, which contain syntactic-semantic conflict, recovery from misinterpretation should be faster when the trial follows an incongruent Stroop trial rather than a congruent Stroop trial. This is because the incongruent Stroop trial is expected to trigger conflict resolution processes, which would be then expected to facilitate resolution of the conflict between syntax and semantics during sentence processing (*if* conflict adaptation is domain-general). Because conflict adaptation is sensitive to decay, the use of eye tracking, rather than keyboard or button responses at the end of the sentence, may help in the detection of these effects by outlining the entire time course of comprehension in a trial.

Figuring out the relationship between conflict monitoring/control and language is important because strengthening these conflict-related systems has potential relevance to therapy for neurological patients with sentence comprehension deficits. If executive functioning were causally involved in language, the implication would be that training patients in executive functioning could improve their ability to process complex language. To date, there exists some evidence that patients also do incremental processing and that a failure to revise an original misinterpretation might underlie some of their errors during sentence comprehension. Meyer and colleagues examined the comprehension of passive sentences by agrammatic aphasics using eye-tracking (Meyer, Mack, & Thompson, 2012). Patients' eye movements showed different patterns on correct versus incorrect response trials. For correct trials, patients appeared to have waited before they fixated on a picture. For incorrect trials, on the other hand, they fixated on the incorrect image immediately following the verb and were unable to repair this mistake. Under the view that executive function is the process by which humans recover from a mistake and adjust behavior dynamically, we would expect that triggering this function would assist in the recovery. This thesis is a first step towards the goal of examining the usefulness of executive function training for language disorders. In order to attain such a goal, we would need to demonstrate conflict adaptation effects first in neurotypical individuals and then in patients. Finally, we would need to extend this knowledge to training or treatment paradigms for patients.

Methods

Participants

Participants included 24 students (3 male, 21 female) at The George Washington University between 18-25 years of age (mean age 18.9). All participants were right-handed, native-English speakers, with visual acuity either normal or corrected-to-normal with contact lenses or glasses. For the purposes of this study, native English speakers were defined as individuals who began learning English before age 6, spoke English in the home growing up, and had at least one native-English speaking parent. Participants were recruited through university psychology courses and were compensated with course credit or $5 cash. They provided consent under a protocol approved by the institutional review board at The George Washington University.

Tasks and Stimuli

The experimental trial sequence contained two types of tasks, Stroop and sentence comprehension. In the Stroop task, participants were shown a word and asked to indicate the ink color of that word by pressing a colored key. In the sentence comprehension task, participants heard a sentence and saw four pictures; they were asked to indicate the image that matched the sentence they heard by pressing a corresponding key. In a congruent Stroop trial, the word shown corresponded to the ink color of the text (e.g. “green” written in green), while in an incongruent trial, the word shown did not correspond to the ink color of the text (e.g. “brown” written in green). Critical sentence trials were presented in the passive voice and described a situation that was either plausible, where the agent and patient filled the expected thematic roles (e.g. “The mouse was chased by the cat”), or implausible and contained syntactic-semantic conflict as the agent and

patient assignments violated their expected thematic roles (e.g. “The cat was chased by the mouse”).

Participants responded using a keyboard. The number-pad keys 4, 5, and 6 were color-coded to correspond to the possible Stroop responses (blue, green, and yellow). Number-pad keys 7, 9, 1, and 3, corresponded to each of the 4 quadrants of the screen where images were displayed during sentence trials (7=top left, 9=top right, 1=bottom left, 3=bottom right). With this keyboard configuration, participants used their right hand to make all responses.

In Stroop trials, one of 6 color words was presented: red, orange, yellow, green, blue, or brown. As only three ink colors were selected for response by keyboard press (blue, green, yellow), Stroop trials contained representational conflict but not response conflict i.e. when a color word conflicted with ink color response (incongruent trial), the color represented by reading that word was not a possible response. For example, when the word “brown” was written in green (the correct response would be green), there was no option for the subject to press a brown key as a response. We manipulated representational rather than response conflict during the Stroop task because this was the type of conflict present during the sentence comprehension task (conflict between two representations or interpretations).

Four images were presented with each critical sentence: target, reversal, and 2 unrelated distractor images (distractor images did not overlap with the target nouns or verb). Sentence trials were counterbalanced such that target images appeared equally often in each of the quadrants of the display. Filler sentence trials displayed a target

image and three foil images that contained the same agent, patient, or verb as the target, or a combination of two.

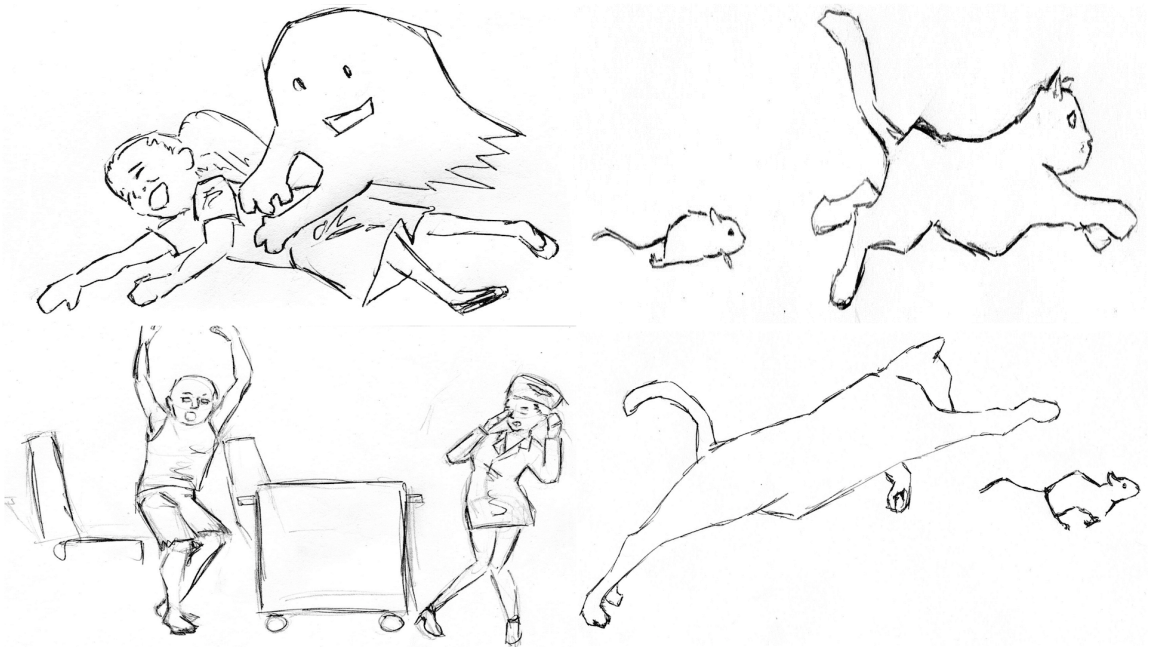

Figure 1: Example for trial "The mouse was chased by the cat." In this example, the target picture is in the bottom right quadrant, the reversal is in the top right quadrant, and the two distractors are in the top and bottom left.

Procedure

Participants were scheduled for 30 minute testing sessions. Participants were seated in front of the eye tracker and the chin rest was adjusted so they could see the entire display screen. The experimenter remained in the room to provide instructions to participants and to calibrate the eye tracking equipment.

Participants were situated in a chinrest 92 cm from a 53 cm x 30 cm Ben-Q Display monitor equipped with Experiment Builder software (version 1.10.1025). The display screen was set at a resolution of 1920x1080, with a viewable portion measuring 28.4cm x 21.3 cm. A host computer was connected to the display monitor and was used

for calibration as well as recording data. An Eyelink-Plus 1000 eye-tracking device was used to record fixations, blinks, and saccades of participants' right eyes (note that participants used both eyes during the experiment). The eye tracker was aligned horizontally with the display monitor. The camera-to-screen distance was 45 cm. The camera tracked eye movements every millisecond (temporal resolution = 1000 Hz).

The timing of each task type was as follows: Stroop trials began with the presentation of a fixation cross, centered on screen, for 500ms. Then, the color word appeared, printed in lowercase letters, for 1000ms in the center of the display screen before a blank screen was displayed for 1000ms. The trial advanced to the blank screen once a response was provided or it timed out, moving to the next trial, if no response was given.

Sentence trials began with the presentation of a fixation cross for 500ms, centered on screen. Then, pictures were displayed in each of four quadrants, while a sentence was played (average of 2462.91ms in length). Sentence length did not vary significantly across conditions. Participants were only able to respond once the sentence finished playing. The pictures remained on screen for 5000ms or until a response was made, before the trial advanced to a blank display screen for 1000ms. The trial timed out and advanced to the next trial if the participant had not made a selection by that time.

Before the experimental block, participants completed practice trials, which followed the same timing sequence as the Stroop and sentence comprehension trials outlined above. The practice trials included 10 Stroop trials, a Stroop baseline block, 10 sentence comprehension trials, and 27 mixed trials (17 Stroop + 10 sentence comprehension). In order to ensure participants' understanding of the tasks, they had to

reach 80% accuracy in Stroop and sentence comprehension blocks to continue onto the Stroop baseline and mixed practice blocks, respectively. If participants did not reach 80% accuracy on either the Stroop or sentence comprehension block, they would repeat that practice block before advancing to the next block. The baseline Stroop block contained 145 trials: 73 congruent and 72 incongruent (36 each of CC, IC, II, and CI. The first trial, which was congruent, was not analyzed). The sentences included in the practice blocks were active sentences. None were included in the actual experiment. All participants completed the same version of the practice.

There were 8 versions of the experimental block. Each of these trial sequence lists contained a total of 120 Stroop trials (60 congruent - C, 60 incongruent - I), 40 passive critical sentence trials (20 plausible - C sentences and 20 implausible - I sentences), and 40 active filler sentence trials. Each participant saw a target sentence and its reversal during the task sequence, so all critical sentence stimuli were seen by all participants, though in a different order. In Stroop trials, the 6 color words (blue, green, yellow, orange, red, brown) were presented 20 times each and the three text colors (blue, green, yellow) were the target response 40 times each, so there was no difference in the number of times a particular Stroop stimulus word appeared or a particular key response was made.

Each trial sequence list was structured such that critical sentences were preceded and followed by Stroop trials, so critical sentence trials never followed one another. There were 40 critical Stroop to Sentence pairs: 10 each in the CC, IC, CI and II conditions. A CC pair would mean that a congruent Stroop trial was followed by a congruent sentence trial; an IC pair would mean that an incongruent Stroop trial was followed by a congruent sentence trial, and so on. The remaining 40 Stroop trials and the

filler sentences were distributed pseudorandomly, resulting in all possible pairs (Stroop to Stroop, Stroop to sentence, sentence to Stroop, sentence to sentence). This meant that participants could not easily predict what the next trial or task was going to be.

We divided incongruent sentences intuitively based on the anticipated level of conflict (e.g., "The matador was killed by the bull" was categorized as low conflict because matadors may both kill and be killed by bulls, while "The nurse was fed by the patient" was categorized as high conflict because nurses typically feed others but are not typically fed by others). Based on this intuitive categorization, we assigned equal numbers of low and high conflict items to the CI and the II conditions within a list such that the sentences in the two conditions did not differ in the overall extent of conflict. To additionally rule out potential differences in conflict among sentences assigned to different conditions, we created 8 lists as mentioned above: any given sentence was preceded by a congruent Stroop trial in half the lists and an incongruent Stroop trial in the other half. By ensuring that an implausible sentence appeared in both CI and II conditions, significant results could not later be attributed to inherent differences between specific sentences. For example, consider the sentence "The cat was chased by the mouse" and its reversal. In List 1, this sentence would be seen in CI condition and its reversal in CC. In List 2, this sentence would appear in CI condition, but its reversal would be seen in IC condition. In list 3, the sentence would be seen in II condition and its reversal in CC. In list 4, the sentence would appear in II condition but its reversal in IC. This way, every possible combination of sentence assignment to condition was covered. In addition, the presentation order of sentences within each list was randomized such that the same sentence did not appear in the same sequence position in different lists (thereby ruling out

unintended order effects). Task order was also varied: lists 5-8 were created by reversing the order of lists 1-4 respectively.

Norming Study

To confirm our intuitions regarding the plausibility or implausibility of the critical sentences, we conducted a survey (see appendix). The survey revealed that items implicitly categorized as incongruent generated greater conflict as intended. Two sentences did not yield a difference score large enough to indicate a significant difference in plausibility; therefore, these items were discarded. Thus, during incongruent sentences, we can expect listeners to experience conflict and look at the more semantically plausible but incorrect picture and subsequently revise that misinterpretation. The central question here is whether the type of the previous Stroop trial would make a difference in the speed of that revision.

Results

Accuracy and Reaction Time (RT)

A mixed effects logistical regression model was used for analysis of accuracy, which is a dichotomous measure (glmer function with the bobyqa optimizer. lme4 version1.1.10 in R, Bates, Maechler, Bolker, & Walker, 2015). The model included a fixed effect of congruency and random intercept and slope by participant. Reaction time was modeled with a mixed effects linear regression (lmer function, lmerTest version 2.0.29). Mean accuracy for the baseline Stroop block was 98.06%. We found a significant effect of congruency for accuracy (estimate=-.96, S.E.=.35, z=-2.71, $p<.01$. Mean accuracy for C trials=98.84%, I trials=97.28%). RT also showed a significant effect of congruency (estimate=28.55, S.E.=4.37, df=23.13, t=6.53, $p<.001$. Mean RT for C trials=530 ms, I trials=558 ms). During the experimental block, mean accuracy on the Stroop trials was 94.10%, and did not vary significantly by congruency ($p>.1$). We also found RTs were slower in the incongruent condition versus the congruent condition (fixed effect of congruency: estimate=36.90, S.E.=6.77, df=69.71, t=5.45, $p<.001$. Mean RT for C trials=619 ms, I trials=657 ms). Together, these results indicate that participants experienced conflict on incongruent Stroop trials.

Accuracy in critical sentence comprehension trials (subjects choosing the correct target image by keyboard press) was high (mean=92.92%). Accuracy in the different conditions (CC, IC, CI and II) was not significantly different. A mixed effects logistical regression model was used and included fixed-effects of previous Stroop trial type and current sentence trial type and their interaction, and random intercepts and random slopes

to best account for variability between participants and items (Barr, Levy, Scheepers, & Tily, 2013). There were no significant effects.

Eye Movements

As we are looking for the adaptation effect from Stroop to Sentence (effect of previous on current), only correct Stroop to correct Sentence trial pairs were analyzed. The analysis period of interest began following the onset of the verb (plus 200ms for eye movement programming). This starting point was selected because the verb is the point at which disambiguation between active and passive sentences occurs within our experimental stimuli. Some verbs were disambiguated as past participles earlier than others (e.g., "bitten" was disambiguated earlier than "chased"), so we began at verb onset to be as inclusive as possible. The window ended at "by": at this point, all sentences were unambiguously passive constructions. Thus, we should expect to see effects of conflict as well as recovery from that conflict within this time window. See Figure 2 for eye movements to target and to reversal images over time and by condition type, where the grey region represents the "verb-ed by" portion of each sentence. No reliable effects were found prior to the "verbed-by" window for either looks to target or looks to reversal.

The analysis window is made up of 100ms bins (following verb onset +200ms) representing looks to target and reversal aggregated by participant and by stimulus, as eye movement data are sategorical. Per Barr (2008), looks to the pictures were empirical-logit transformed and then subjected to mixed-effects regression analysis. This mixed-effects model contained the fixed-effects of previous Stroop and current sentence trial types and their interaction, and all applicable random effects. For looks to the target picture, there were no significant effects of previous or current trial type and no interaction ($p \geq .1$, See

Table 1). For looks to the reversal picture, there was a significant interaction consistent with conflict adaptation: for incongruent sentences, which induce syntactic-semantic conflict, there was a significant effect of previous Stroop trial type; for congruent sentences, which do not involve conflict, there was no such effect. See Figure 3 for mean proportion of looks to reversal by condition type. Looks to the reversal were lower in the II than in the CI condition (Participants analysis: $t(23)=-2.52$, $p<.05$. Items analysis: $t(17)=-2.77$, $p<.05$) but not significantly different between the CC and IC conditions (Participants analysis: $t(23)=1.68$, $p>.1$. Items analysis: $t(17)=.81$, $p>.4$) (See Table 2).

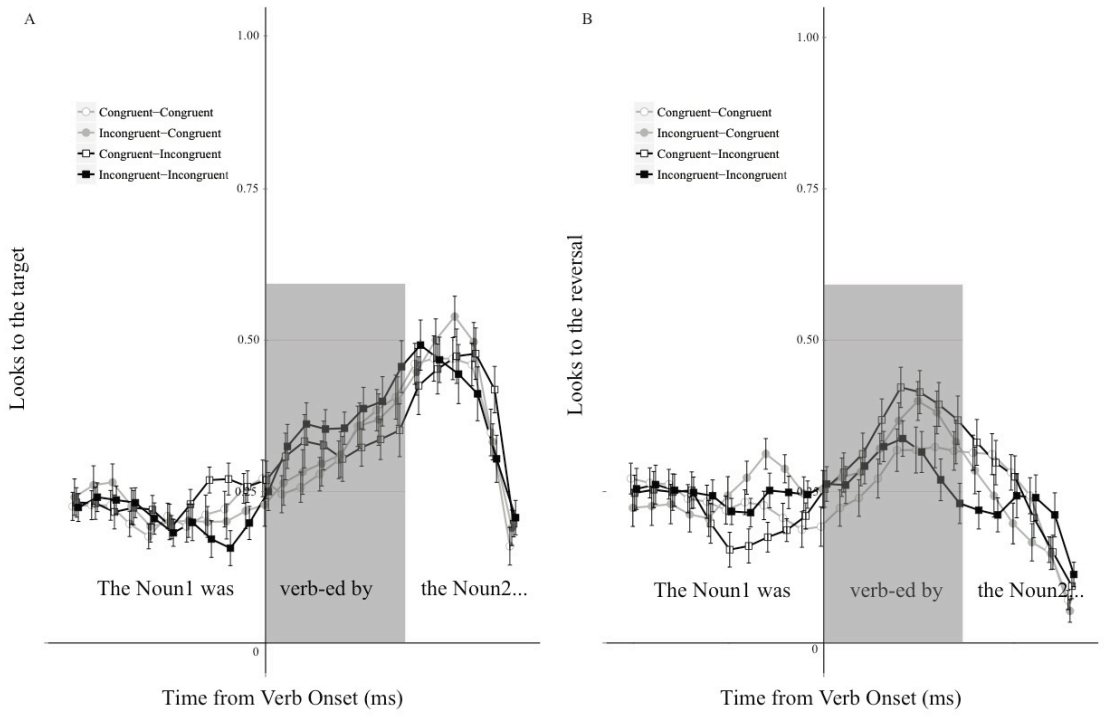

Figure 2: Proportion of looks to the target (A) and the reversal (B) picture in different conditions

Table 1: Mixed models of the likelihood of target looks in the "verb-ed by" window

	Estimate	S.E.	t	p
Target looks aggregated by participant				
Fixed effects				
Intercept	-1.07	.11	-9.63	<.001
Previous trial type	.13	.28	.48	.64
Current trial type	.10	.25	.41	.69
Interaction	.69	.41	1.67	.11
Target looks aggregated by item				
Fixed effects				
Intercept	-1.16	.23	-5.10	<.001
Previous trial type	.28	.23	1.19	.24
Current trial type	.03	.45	.06	.96
Interaction	.31	.47	.66	.51

Table 2: Mixed models of the likelihood of reversal looks in the "verb-ed by" window

	Estimate	S.E.	t	p
Reversal looks aggregated by participant				
Fixed effects				
Intercept	-1.06	.09	-12.46	<.001
Previous trial type	-.04	.16	-.27	.79
Current trial type	.20	.20	1.00	.33
Interaction	-.91	.31	-2.92	<.01
Reversal looks aggregated by item				
Fixed effects				
Intercept	-1.42	.24	-5.93	<.001
Previous trial type	-.32	.23	-1.43	.16
Current trial type	.11	.48	.22	.82
Interaction	-1.16	.45	-2.55	.02

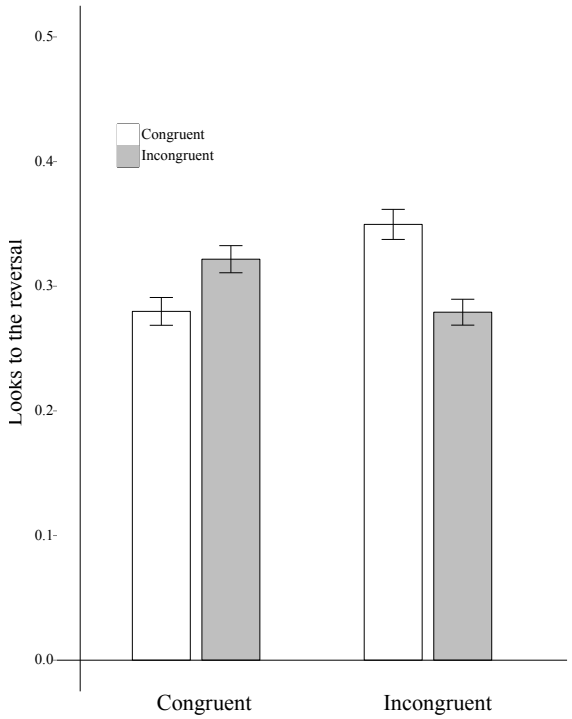

Figure 3: Proportion of looks to the reversal picture in different conditions within the "verb-ed by" window.

Individual Differences

Individual differences in subjects were considered to see if the impact of the conflict adaptation effect differed based on subjects' performance on the Stroop baseline block. The Stroop effect was calculated by subtracting the reaction time (RT) for congruent Stroop trials from the RT for incongruent Stroop trials to see how much longer RTs were on incongruent trials. The eye movement suppression effect was based on the proportion of looks to the reversal image in the II minus the CI condition. We analyzed the correlation between the Stroop and eye movement suppression effects with the skipped Pearson correlation analysis using the robust correlation toolbox in Matlab (Pernet, Wilcox, & Rousselet, 2013). Results revealed a significant negative correlation

indicating that participants who showed a larger Stroop effect showed greater suppression of the reversal image in II vs. CI conditions (See Figure 4, r=-.38, CI=[-.63 -.15]).

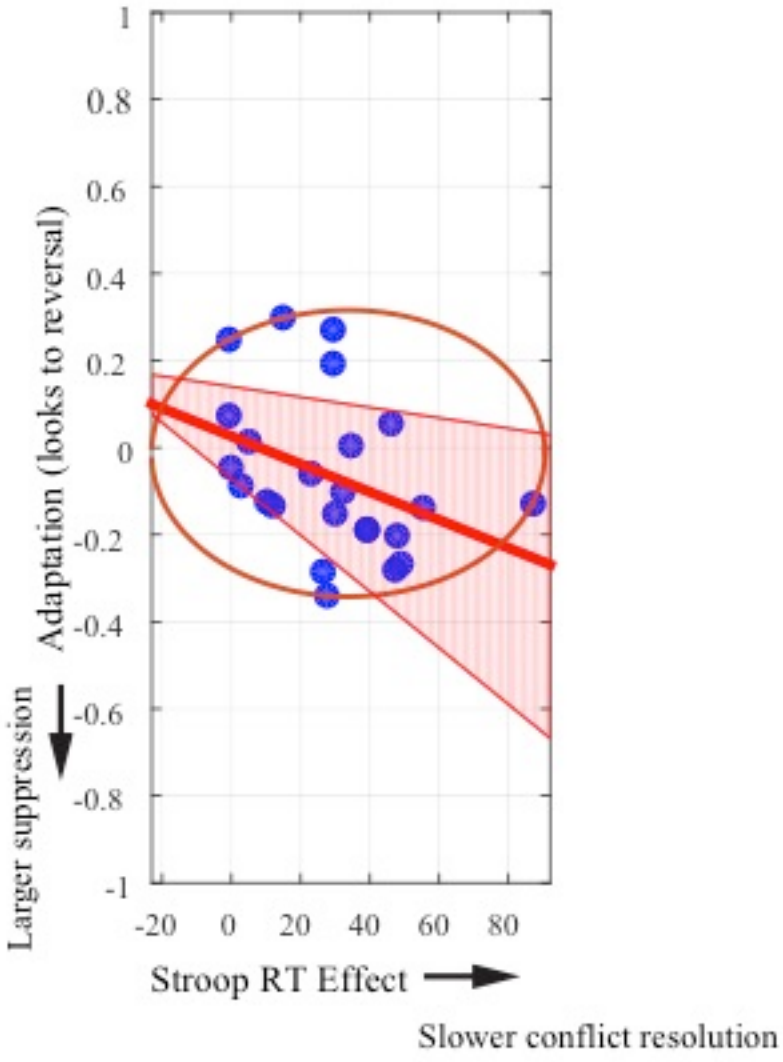

Figure 4: Significant negative correlation between different participants' Stroop effects and the suppression of their looks to the reversal during incongruent sentences.

Potential Source of the Individual Differences

We then explored possible mechanisms by which those participants showing a greater Stroop effect at baseline achieved greater adaptation. We hypothesized that they were particularly susceptible to using semantic cues in both the Stroop and the sentence comprehension tasks. In this case, the observed adaptation in these participants might have been because they had to suppress the semantic information (the meaning of the color word) during incongruent Stroop trials and this suppression extended to the following sentence trial and facilitated resolution of syntactic-semantic conflict. To further explore the source of the individual differences in conflict adaptation reported above, agent/patient ratings from the norming study (see appendix) were used to split items into high-conflict and low-conflict. High-conflict items (e.g. "The doctor was treated by someone") received a high agentive rating as doctors are more likely to treat than be treated and low-conflict items (e.g. "The cat was chased by another animal") received a low agentive score as cats are likely both to chase and to be chased. If participants are differentially guided by semantic information, then those with an above-median Stroop effect (those who have comparatively more trouble suppressing semantic information inherent to the color word during the Stroop task) should be impacted by the semantic cues in the high-conflict items and show more looks to the reversal as compared to the low-conflict sentences.

A mixed-effects regression analysis was used where looks to reversal were transformed by the empirical-logit function. The "verbed by" window was again selected as the window of interest. The model included fixed effects of previous trial type and item type (high or low-conflict as described above), and all applicable random effects.

There was a significant effect of previous trial type (which is expected because these participants were the ones who showed larger conflict adaptation effects). There was also a significant effect of item type (refer to Figure 5). High-conflict items showed increased looks to the reversal for these participants, consistent with our prediction that these participants would show sensitivity to the strength of the semantic cues.

For comparison, we conducted the same analysis as above with participants showing a below median Stroop effect. In this set of participants, we did not detect an effect of item type ($p>.1$. See Table 3). Thus, only those participants showing a greater Stroop effect at baseline showed a difference in processing high-conflict and low-conflict sentence trials, which may indicate that they relied more on semantics and benefitted more from semantic suppression. This proposal should be considered tentative however: when considering both types of subjects together, the two-way interaction between participant type and item type was not significant ($p=.18$). Future experiments are needed to corroborate this speculative analysis.

Table 3: Mixed models of the likelihood of reversal looks to high and low conflict items in participants who showed large and small Stroop effects

	Estimate	S.E.	t	p
Large Stroop effect				
Fixed effects				
Intercept	-1.62	.24	-6.77	<.001
Previous trial type	-1.49	.36	-4.14	<.01
Item type	1.00	.44	2.28	<.05
Interaction	-.19	1.06	-.18	.86
Small Stroop effect				
Fixed effects				
Intercept	-1.67	.25	-6.78	<.001
Previous trial type	-.27	.42	-.63	.54
Item type	.15	.45	.33	.75
Interaction	.46	.65	.71	.49

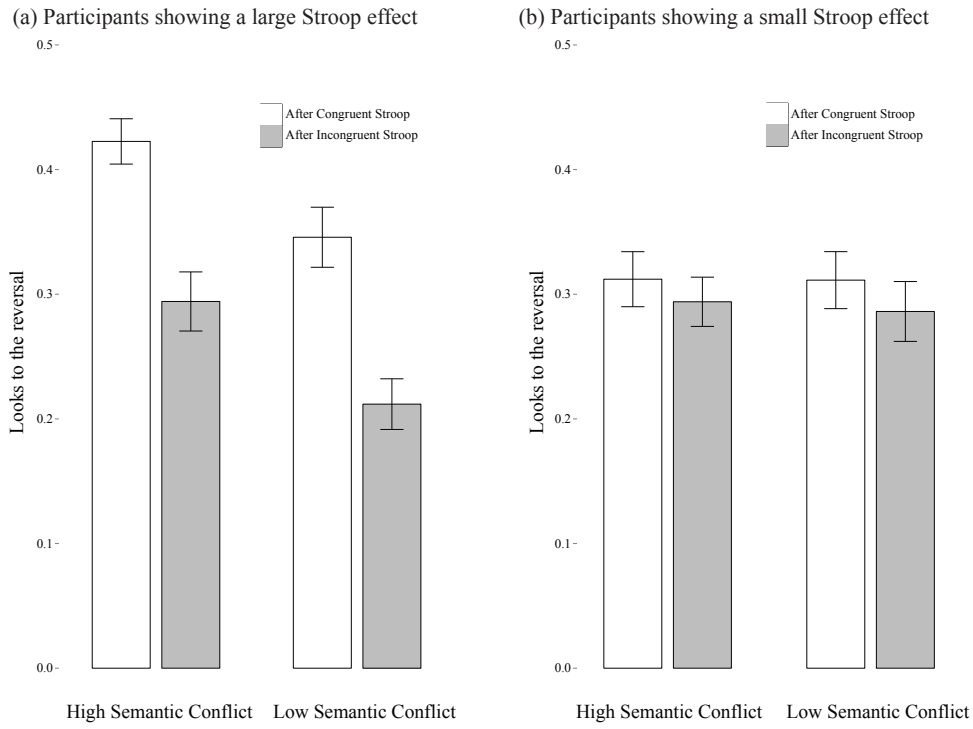

Figure 5: Proportion of looks to the reversal picture for different item types within the "verb-ed by" window

Discussion

Our aim was to explore the presence and manifestation of the conflict adaptation effect between Stroop and the comprehension of sentences containing syntactic-semantic conflict. We used an eyetracker because adaptation effects are sensitive to decay and keyboard press reaction time does not capture the full picture of on-line sentence processing. A norming study was additionally conducted to measure the level of conflict produced by our experimental stimuli in terms of the specific scenarios described and the agentive ratings of the noun entities described in the sentences. We first discuss the norming results and then the conflict adaptation results.

The results of the norming study indicated that the experimental stimuli that were analyzed generated sufficient conflict and were assigned to the appropriate plausible or implausible condition. See appendix for full discussion of the norming study.

Turning to the conflict adaptation results: If the previous Stroop trial were to have an effect on the current sentence trial (by triggering conflict resolution processes that remain active during comprehension of the sentence trial), then we would expect an implausible sentence following an incongruent Stroop (II) to have a faster resolution of conflict than one that follows a congruent Stroop (CI). If this is the case, looks to the target image should be achieved faster for II than CI sentences, and there should be fewer looks to the reversal for II than CI sentences. This effect would not be predicted for CC versus IC sentences, as they do not contain conflict, making adaptation processes inapplicable. If the conflict adaptation effect were detected in the "verbed by" time window of relevance, there would be an interaction among condition types whereby there

would be a significant difference in the time course of looks to target/reversal between II and CI but not between CC and IC.

Our eye-movement results corroborate the presence of conflict adaptation. The analyses revealed significantly reduced looks to the reversal (and semantically more plausible) image for incongruent sentences that followed an incongruent Stroop relative to those that followed a congruent Stroop. These findings support the notion that the conflict adaptation effect can be present across tasks and that processes activated to resolve conflict on the previous trial influence performance on a subsequent trial containing conflict. Our study contributes to the literature to show the extent of adaptation; in this case, it is purported to exist across a task containing no syntactic conflict (as in the Stroop) to a task containing linguistic conflict of a syntactic nature (comprehension of syntactic-semantic conflict sentences). Previous imaging evidence provides correlational evidence as to the involvement of the left inferior frontal gyrus (known for its role in processing conflict) in the processing of sentences containing conflict. Our study provides *causal* evidence that supports domain-general accounts for the role of executive function across task types. Even across task types, cognitive control mechanisms remain active and are used after being triggered by the presence of conflict.

Cognitive control mechanisms may involve suppression of task-irrelevant information, increased attention to task-relevant information, or a combination of the two. Because we found significant results for the suppression of looks to the reversal but did not find significant results for increased looks to the target, it appears that our results support the hypothesis that competing, task-irrelevant information (in this case semantic

information about the entities described in the sentence) is suppressed in order to facilitate recovery from misinterpretation.

Subjects did not have increased looks to the target image even though they demonstrated decreased looks to the reversal during II trials. In addition to indicating that task-relevant information may not have been selectively highlighted, this behavior indicates that subjects were looking at the other distractor images when they were not looking at the reversal. One hypothesis to explain this phenomenon is that subjects may have been prepared with the correct keyboard response during this time but were waiting for the sentence to finish before being able to log a response. In this case, they may have taken the time to look at the other distractor images, in a sense "double-checking" their intended response rather than still deciding between target and reversal, as the reversal had already been ruled out.

Individual differences were explored to investigate the mechanisms by which participants resolved conflict and to investigate who is more likely to benefit from conflict adaptation. Our exploration of individual differences was guided by the hypothesis that those who showed a larger Stroop effect are more impacted by semantics. For the Stroop task, greater reliance on semantics would lead to difficulty suppressing the word's meaning. For the sentence comprehension task, a greater reliance on semantic cues would create difficulty in suppressing looks to the reversal in implausible sentences, as the reversal images depict the more semantically plausible situations. Thus, our explanation of the individual differences is as follows: Some participants tend to be more semantically guided. These participants suppress semantics to perform the Stroop task. The semantic suppression mechanisms triggered during an incongruent Stroop trial

remain active during a subsequent incongruent sentence trial, thereby leading to conflict adaptation.

Our stimulus items differed in the amount of conflict they induced and this difference was used to explore the above semantic suppression hypothesis. The norming study was used to classify stimuli into high and low conflict sentences based on agentive ratings (an entity that can fill either the agent or patient role for a given verb creates less conflict in the patient position than an agent that typically only fills the agentive role, but is assigned to the patient role). High-conflict sentences contain stronger semantic cues (and may be more misleading for those relying more on semantics to complete the comprehension task) than low-conflict sentences. Results revealed that participants with a greater Stroop effect showed more looks to the semantically plausible but incorrect image in high-conflict than low-conflict sentences. That is, based on their performance on Stroop tasks at baseline, these subjects were hypothesized to be more greatly impacted by semantic information and indeed were found to be more hampered by the high-conflict sentences. Those who showed less of a Stroop effect at baseline did not show differences in performance on high- or low-conflict sentence trials, suggesting either that they routinely rely on syntax rather than semantics for sentence interpretation or that they were successful at suppressing semantics within the context of this experiment. Future research could explore whether different individuals complete Stroop and sentence comprehension tasks using different pathways or mechanisms.

One limitation of the study is that all filler sentences were presented in the active voice and did not include “was”. In that case, participants may have been sensitive to the potential for conflict appearing at the point of “was verbed by” rather than the “verbed by”

region since no sentences followed the structure “The agent was verb-ing the patient”. It should also be noted that there were no lines explicitly demarcating the four image choices during sentence comprehension trials. The presence of lines may have helped participants more clearly distinguish which image they were looking to. However, the spatial resolution of the camera was sufficient to obtain accurate data regarding which quadrant a participant was viewing at a given time regardless of the presence of lines. An additional consideration is the potential for minimal overlap of some distractor images with their corresponding target/reversal pair; that is, some nouns may have been identified by participants on the basis of gender (e.g. “doctor” seen as “woman”). Still, such a distractor would have overlapped only by one dimension (one of the nouns), while the reversal, which overlapped by three dimensions (both nouns and the verb), still generated greater conflict and competition with the target.

Though this study explored executive function, there are always other possible factors that could have impacted a subject’s performance on this task. A person’s decision making, as demonstrated by their eye movements, could by impacted by attention, motivation, or personality characteristics. Regardless, we compare performance relatively.

Future directions should aim to replicate these results in older, neurotypical adults as well as explore the adaptation effect in patient populations, because there could be important clinical implications. With this thesis, we have shown conflict adaptation effects in neurotypical individuals and argued for a domain-general account of executive functioning measures. If this finding were true in patients as well, then this relationship between language and executive function could be exploited for therapy techniques

whereby strengthening a patient's executive functioning abilities could lead to improvements in sentence comprehension abilities. Further, we discovered that individuals who relied more on semantic cues benefitted more from a previous incongruent Stroop trial. This "semantic route" to comprehension is a common pattern seen in some aphasic patients (Saffran, Schwartz, & Linebarger, 1998). Such individuals might specifically benefit from triggering conflict resolution using Stroop or other general cognitive control tasks. The sustained activation of these mechanisms might then assist the patients to suppress semantics and use syntax instead. Of course, this assumes that the patients have usable syntactic and executive function abilities. It seems unlikely that this mechanism would work for patients who cannot interpret sentences without using semantic cues or for those who have no remaining executive function abilities following brain damage. While acknowledging these potential limitations, this study is the first to show that cognitive control speeds up suppression of task-irrelevant information during the comprehension of sentences containing syntactic-semantic conflict. Application of these findings to aphasic patients, be it even a small subset of these individuals, would be a novel approach to therapy that can complement existing techniques.

References

Altmann, G.T.M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57*(4), 502-518. doi:10.1016/j.jml.2006.12.004

Barr, D.J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language. 59*(4), 457-474. doi: 10.1016/j.jml.2007.09.002

Barr, D.J., Levy, R., Scheepers, C., & Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keeping it maximal. *Journal of Memory and Language, 68*(3), 255-258. doi: 10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01

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. doi: 10.1037//0033-295X.I08.3.624

Constantinidou, F., Wertheimer, J.C., Tsanadis, J., Evans, C., & Paul, D.R. (2012). Assessment of executive function in brain injury: Collaboration between speech-language pathology and neuropsychology for an integrative neuropsychological perspective. *Brain Injury, 26*(13-14), 1549-1563. doi: 10.3109/02699052.2012.698786

Egner, T. (2008). Multiple conflict-driven control mechanisms in the human brain.

Trends in Cognitive Sciences, 12(10), 374-380. doi: 10.1016/j.tics.2008.07.001

Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience, 8*(12), 1784-1790. doi: 10.1038/nn1594

Freitas, A.L., Bahar, M., Yang, S., & Banai, R. (2007). Contextual adjustments in cognitive control across tasks. *Psychological Science, 8*(12), 1040-1043. doi: 10.1111/j.1467-9280.2007.02022.x

Freitas, A.L., & Clark, S.L. (2015). Generality and specificity in cognitive control: conflict adaptation within and across selective-attention tasks but not across selective-attention and Simon tasks. *Psychological Research, 79* (1), 143-162. doi: 10.1007/s00426-014-0540-1

Hyafil, A., Summerfield, C., Koechlin, E. (2009). Two mechanisms for task switching in the prefrontal cortex. *The Journal of Neuroscience, 29*(16), 5135-5142. doi: 10.1523/JNEUROSCI.2828-08.2009

January, D., Trueswell, J.C., & Thompson-Schill, S.L. (2009). Co-localization of Stroop and syntactic ambiguity resolution in Broca's area: Implications for the neural basis of sentence processing. *Journal of Cognitive Neuroscience, 21*(12), 2434-2444. doi: 10.1162/jocn.2008.21179

Kan, I.P., Teubner-Rhodes, S., Drummey, A.B., Nutile, L., Krupa, L., & Novick, J.M. (2013). To adapt or not to adapt: The question of domain-general cognitive control. *Cognition, 129*(3)*,* 637-651. doi: 10.1016/j.cognition.2013.09.001

Kerns, J.G., Cohen, J.D., MaDonald, A.W., Cho, R.Y., Stenger, A., & Carter, C.S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science,*

303(5660), 1023-1026. doi: 10.1126/science.1089910

Kornblum, S., Hasbroucq, T, & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility – A model and taxonomy. *Psychological Review, 97*(2), 253-270. doi: 10.1037/0033-295X.97.2.253

MacDonald, A.W., Cohen, J.D., Stenger, V.W., & Carter, C.S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulated cortex in cognitive control. *Science, 288*(5472), 1835-1838. doi: 10.1126/science.288.5472.1835

Malaia, E., & Newman, S. (2015). Neural bases of syntax–semantics interface processing. *Cognitive Neurodynamics, 9*(3), 317-329. doi: 10.1007/s11571-015-9328-2

Mayr, U., Awh, E., & Laurey, P. (2003). Conflict adaptation effects in the absence of executive control. *Nature Neuroscience, 6*(5), 450-452. doi:10.1038/nn1051

Meyer, A.M., Mack, J.E., & Thompson, C.K. (2012). Tracking passive sentence comprehension in agrammatic aphasia. *Journal of Neurolinguistics, 25*(1), 31-43. doi: 10.1016/j.jneuroling.2011.08.001

Miller, E.K., & Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*, 167-202. doi:10.1146/annurev.neuro.24.1.167

Novick, J.M., Thompson-Schill, S.L., & Trueswell, J.C. (2008). Putting lexical constraints in context into the visual-world paradigm. *Cognition, 107*(3), 850-903. doi: 10.1016/j.cognition.2007.12.011

Pernet, C.R., Wilcox, R., & Rousselet, G.A. (2013). Robust correlation analyses: false positive and power validation using a new open source Matlab toolbox. *Frontiers in Psychology. 3*(606), 1-18. doi: 10.3389/fpsyg.2012.00606

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372-422. doi: 10.1037/0033-2909.124.3.372

Saffran, E.M., Schwartz, M.F., & Linebarger, M.C. (1998). Semantic influences on thematic role assignment: Evidence from normals and aphasics. *Brain and Language*, *62*(2), 255-297. doi: 10.1006/brln.1997.1918

Tanenhaus, M.K., & Trueswell, J.C. (2006). Eye movements and spoken language comprehension. In M.J. Traxler & M.A. Gernsbacher (Eds.) *Handbook of psycholinguistics* (863-900). (2nd ed.). London, UK: Academic Press.

Ullsperger, M., Bylsma, L.M., & Botvinick, M.M. (2005). The conflict adaptation effect: It's not just priming. *Cognitive, Affective, & Behavioral Neuroscience, 5*(4), 467-472. doi: 10.3758/CABN.5.4.467

Ye, Z., & Zhou, X. (2008). Involvement of cognitive control in sentence comprehension: Evidence from ERPs. *Brain Research, 1203*, 103-115. doi: 10.1016/j.brainres.2008.01.090

Ye, Z., & Zhou, X. (2009). Conflict control during sentence comprehension: fMRI evidence. *NeuroImage, 48*(1), 280-290. doi: 10.1016/j.neuroimage.2009.06.032

Appendix

Survey participants included 23 subjects (16 females, 7 males) aged 18-40 (average age 25) fitting the qualification criteria for native English speakers described earlier. None of the subjects had neurocognitive disorders. None of the participants of the norming study participated in the eye tracker experiment, though two were involved in a previous iteration of this experiment. The survey was administered through Surveymonkey (https://www.surveymonkey.com). It was divided into 3 sections: plausibility ratings, agent/patient ratings, and noun prominence ratings, each with a 6-point Likert-style response scale.

In the plausibility ratings section, participants were asked: "How likely is the scenario described by the sentence?" -5 (highly unlikely), -3(moderately unlikely), -1(slightly unlikely), +1(slightly likely), +3(moderately likely), +5(highly likely). Order of presentation of items was counterbalanced such that a scenario and its reversed version appeared sufficiently far apart (mean distance 20). All but two scenario pairs were found to have a significant rating difference between the sentence that we had intuitively categorized as plausible and the reversed sentence that we had categorized as implausible (mean difference=7.90; $t(17)=16.39$, $p<.001$). Two sentence pairs were found to have a rating difference <1, indicating that neither sentence was clearly rated more plausible than the other. These sentence pairs were, "The bull was killed by the matador"/"The matador was killed by the bull" and "The dragon was killed by the prince"/"The prince was killed by the dragon". These sentences were removed from analysis.

In the agent/patient ratings section, participants were asked to consider each pair of scenarios to determine which was more likely, active or passive, by answering the

question: Which scenario is more likely? “Noun verbed someone/an(other) animal.” “Noun was verbed by someone/an(other) animal.” For example, participants would see sentence pairs like the following: “The mouse chased another animal. The mouse was chased by another animal.” “The man kicked an animal. The man was kicked by an animal.” or “The doctor treated someone. The doctor was treated by someone.” and rate them on a scale -5 (Sentence 1 is highly more likely than Sentence 2)... +5 (Sentence 2 is highly more likely than Sentence 1). As before, we ensured that a scenario and its reversed version appeared sufficiently far apart (mean distance 20). Negative numbers indicate an agent bias and positive numbers indicate a patient bias. Results showed that the entities assigned to patient roles in our incongruent sentence stimuli (e.g. “cat” in “The cat was chased by the mouse”) were rated as more likely to be the agents of the verbs used in those sentences (thereby leading to conflict) (mean agent rating =-3.57, $t(17)=-11.33$, $p<.001$). In contrast, the entities assigned to patient roles in the congruent sentence stimuli (e.g. “mouse” in “The mouse was chased by the cat”) were rated as more likely to be the patients of the verbs used in the sentences (thereby leading to no conflict) (mean patient rating =2.92, $t(17)=7.74$, $p<.001$).

In the noun prominence rating section, participants rated the saliency of the nouns pictured in the plausible and implausible images by responding: “Rate whether the entity listed below is more visible in the left or right picture”: -5 (Highly more visible in left picture)... +5 (Highly more visible in right picture). The plausible image appeared on the left for half of the items. These ratings were obtained to explore potential visual confounds. Results indicated that the entities assigned to patient roles in implausible sentence stimuli (e.g. “cat”) were rated as more visible when they were agents of the

depicted action and the entities assigned to patient roles in plausible sentence stimuli (e.g. "mouse") were rated as more visible when they were patients (mean agent picture preference for incongruent sentence subjects =2.71, $t(17)=-9.91$, $p<.001$; mean patient picture preference for congruent sentence subjects=1.38, $t(17)=4.25$, $p<.001$ respectively).

To summarize, all three sections of the norming survey confirmed our prediction that incongruent sentences were less plausible and would give rise to conflict. Survey participants gave higher plausibility ratings to sentences that we categorized as congruent than to sentences that we categorized as incongruent. They rated the subjects of congruent sentences as more likely to be patients of the verbs (consistent with a passive syntactic structure) but the subjects of incongruent sentences as more likely to be agents of the verbs (conflicting with a passive syntactic structure). For example, participants may have been temporarily misled as they processed a passive sentence, "The doctor was treated by the patient" and assumed "doctor" to fill an agentive role due to this bias towards doctors treating others rather than being treated by others. Finally, image saliency also reflected plausibility in that entities expected to take an agentive role were considered more visible when they filled the agent role and entities expected to take a patient role were considered more visible when they filled the patient role. Listeners experience more conflict on incongruent sentences, as expected.