Psychological Topics 23 (2014), 1, 167-187

Review article – UDC – 616.8
159.95
159.953

Neuroimaging Evidence of Comprehension Monitoring

Linda Baker, Alisa Zeliger-Kandasamy, Laura U. DeWyngaert

University of Maryland, Baltimore County, USA

Abstract

The purpose of this article is to synthesize the emerging neuroimaging literature that reveals how the brain responds when readers and listeners encounter texts that demand monitoring of their ongoing comprehension processes. Much of this research has been undertaken by cognitive scientists who do not frame their work in metacognitive terms, and therefore it is less likely to be familiar to psychologists who study metacognition in educational contexts. The important role of metacognition in the development and use of academic skills is widely recognized. Metacognition is typically defined as the awareness and control of one's own cognitive processes. In the domain of reading, the most important metacognitive skill is comprehension monitoring, the evaluation and regulation of comprehension. Readers who monitor their understanding realize when they have encountered difficulty making sense of the text, and they apply error correction procedures to attempt to resolve the difficulty. Metacognition depends on executive control skills that continue to develop into early adulthood, in parallel with the maturation of the executive control regions of the prefrontal cortex. Functional magnetic resonance imaging (fMRI) and event-related potentials (ERP) have been used for some time to study neural correlates of basic reading processes such as word identification, but it is only within recent years that researchers have turned to the higher-level processes of text comprehension. The article describes illustrative studies that reveal changes in neural activity when adults apply lexical, syntactic, or semantic standards to evaluate their understanding.

Keywords: metacognition, comprehension monitoring, neuroimaging, reading, comprehension

Introduction

The important role of metacognition in the development and use of academic skills is widely recognized in research and practice (Baker, 2011; Hacker, Dunlosky, & Graesser, 2009). Metacognition is typically defined as the awareness and control of one's own cognitive processes. In the domain of reading, the most important metacognitive skill is comprehension monitoring, the evaluation and regulation of comprehension. Readers who monitor their understanding realize

✉ Linda Baker, Department of Psychology, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA. E-mail: *baker@umbc.edu*

when they have encountered difficulty making sense of the text, and they apply error correction procedures to attempt to resolve the difficulty. Research has documented that although even young children have some capability in monitoring their comprehension, these executive control skills continue to develop into the adult years (Baker & Beall, 2008).

Neuroscience research provides intriguing evidence that developmental changes in metacognition may be related to maturation of the prefrontal cortex, the portion of the brain involved in executive function. Executive function includes processes typically regarded as metacognitive in nature, such as planning, monitoring, and error correction and detection.

Functional magnetic resonance imaging (fMRI) and event-related potentials (ERP) have been used for some time to study neural correlates of basic reading processes (e.g., word identification), but it is only within recent years that researchers have begun to study higher-level comprehension processes that provide evidence of comprehension monitoring (Ferstl, Neumann, Bogler, & von Cramon, 2008; Landi, Frost, Mencl, Sandak, & Pugh, 2013; Perfetti & Bolger, 2004). These processes include generating inferences to bridge coherence gaps in a story, detecting violations of world knowledge, and coping with referential and syntactic ambiguities. The purpose of this article is to synthesize the emerging literature that reveals how the brain responds when adult readers and listeners encounter texts that challenge their ongoing comprehension processes. This research has not yet been situated within a metacognitive framework, and therefore it is less likely to be familiar to psychologists who study metacognition in educational contexts. Those neuroimaging studies that focus on metacognition tend to address memory of discrete items rather than text comprehension (Shimamura, 2008) and are beyond the scope of this review.

The review is organized as follows. For the benefit of readers who are not highly familiar with cognitive neuroscience, we open with an overview of brain regions involved in language comprehension and executive control, followed by a brief description of fMRI and ERP approaches. The review of empirical studies is organized primarily in terms of the standards that readers use to evaluate their comprehension. As described by Baker (1985), readers use standards at the lexical, syntactic, and semantic levels to evaluate their understanding. The semantic level includes external consistency, internal consistency, and propositional cohesiveness. Because the higher-level processes of comprehension are similar in reading and listening, and the neuroimaging results bear this out, we include studies of both types in this review. Subsequent sections examine individual differences and the neural responses to using reading strategies that require metacognitive processing of text. We conclude with a summary and discussion of future directions.

Brain Regions Involved in Comprehension and Executive Control

Comprehension

The extended language network (ELN) (Ferstl et al., 2008) is a term used to describe the interconnected cortical pathways that are involved in language comprehension. Through neuroimaging, researchers have been able to identify parts of the brain that are involved in the comprehension of written and spoken language. The left hemisphere plays a major role, as we have known for some time, but neuroimaging research has revealed right hemisphere involvement that is not as fully understood.

The dorsal, ventral, and anterior systems of the brain are all important to semantic processing. The dorsal, or temporoparietal, system plays a role in semantic analysis, and with mapping visual aspects of print and binding phonological and semantic representations together. The dorsal system includes the angular gyrus (AG), the supramarginal gyrus (SMG), and the posterior aspect of the superior temporal gyrus (STG), or Wernicke's area. The ventral system is particularly attuned to linguistic stimuli. It includes the left inferior occipitotemporal/fusiform region (OT), the middle temporal gyrus (MTG), and the inferior temporal gyrus (ITG). The anterior system plays a role in semantic retrieval and includes the inferior frontal gyrus (IFG), and insular and opercular regions. Broca's area is located within the IFG.

Many neuroimaging studies have been conducted at the word level, fewer at the sentence level, and a small number conducted on comprehension of longer texts. Landi et al. (2013) suggested that many of the same regions involved in single word and sentence processing are also involved in comprehension of longer texts, including the left IFG. However, differences in brain activation are evident when processing longer texts, particularly in the right hemisphere. Longer texts elicit greater right hemisphere and frontal lobe activation, including greater activation of the right precentral sulcus relative to the left and increased activation in the right superior frontal gyrus (SFG) and medial frontal gyrus (MFG). Ferstl et al.'s (2008) meta-analysis included 23 neuroimaging studies that used connected discourse. The research showed that longer passages activate the anterior temporal lobes (aTL), which contribute to sentence comprehension and are linked to syntactic, semantic and episodic memory processes that integrate to give language meaning. The lateral prefrontal cortex (including the frontal operculum) is involved with processing of sentence context. The AG, the posterior parietal cortex (PPC) and the precuneus are associated with the manipulation of mental models and play a role in relating text to prior knowledge. The left IFG is associated with semantic analysis and the right inferior parietal lobes, the right inferior and middle temporal regions are involved in inference making.

Executive Function

Executive functions, or higher-order cognitive abilities involved in goal directed behavior, develop throughout childhood into early adulthood. Executive functioning is highly related to the academic skills of reading and math (Houde, Rossi, Lubin, & Joliot, 2010). Developmental increases in executive function correlate with increased myelination and synaptogenesis of the frontal brain regions (Klingsberg, Vaidya, Gabrieli, Moseley, & Heduhus, 1999). Maturation of the prefrontal lobes begins early in childhood and continues throughout adolescence. Fernandez-Duque, Baird, and Posner (2000) were among the first to link the executive control regions of the brain with metacognitive processes. Executive functions are centered in the prefrontal cortex (PFC), which is made up of the dorsolateral prefrontal cortex (dlPFC), the ventromedial prefrontal cortex (vmPFC) and the orbitofrontal cortex (OFC) (Ardila, 2013). The angular cingulate plays a particularly important role in metacognitive tasks, including conflict monitoring and resolution and error detection, all of which are involved in comprehension monitoring. Because maturation of the prefrontal lobes is not complete until late adolescence or early adulthood, it is easy to understand why lapses in metacognitive control during reading are still apparent in college students (Baker, 1989).

Neuroimaging Methods Used in Research on Comprehension

Functional magnetic resonance imaging (fMRI) and electrophysiological recording of event-related potentials (ERP) are two common neuroimaging approaches in language research. Because they are the two that are used in the studies discussed in this article, we provide brief descriptions.

Functional Magnetic Resonance Imaging (fMRI)

fMRI is a non-invasive technique used to measure and map neural activity that became available to researchers in the 1990s. It works by detecting changes in blood flow within the brain through powerful magnetic and radio waves. Unlike MRI, which images brain structure, fMRI provides information about how the brain functions, such as when participants read or listen to passages. The primary form of fMRI uses the Blood-Oxygen-Level Dependent (BOLD) contrast, which detects the amount of oxygen present in hemoglobin molecules. Deoxygenated hemoglobin is more magnetic than oxygenated hemoglobin, so when neurons in an area of the brain become active, the oxygenated blood flows to the activated areas. The oxygenated blood flow then produces a clear magnetic resonance signal, which can be mapped to show neural activity in specific brain regions.

Event-Related Potential (ERP)

ERP is the electrophysiological response of the brain to specific stimuli, measured with electroencephalography (EEG). EEG is a non-invasive procedure that measures electrical activity of the brain using electrodes placed on the scalp. The technology became widely used by cognitive neuroscientists in the 1980s. ERP waveforms consist of a series of positive and negative voltage deflections, typically referred to by a letter indicating polarity (N or P for negative/positive) followed by a number indicating the time in milliseconds after stimulus presentation. ERP shifts that are important in comprehension research are the N400, the P600, and the Nref. The N400 response is apparent when readers or listeners are presented with words or phrases that do not fit well semantically with the surrounding context. The P600 occurs in response to syntactic processing difficulties and where conflicts exist between semantic and syntactic information. The Nref is a sustained negative shift from the brain's frontal areas, which occurs when readers encounter ambiguous referents. All of these waveforms are indicative of ongoing monitoring of one's comprehension processes.

Comparison of fMRI and ERP Methods

ERP provides excellent temporal resolution, meaning one can determine precisely when the response occurs relative to stimulus onset. However, it does not provide precise information about where in the brain the activity is taking place. In contrast, fMRI has good spatial resolution because it measures a hemodynamic (blood flow) response rather than an electrical response. Temporal and spatial resolution are both important in research on monitoring of understanding because error detection and correction procedures occur at different points in the comprehension process and in different regions of the brain. It would be more informative if researchers were to use both ERP and fMRI (Perfetti & Bolger, 2004), but that is not common.

Neuroimaging Evidence for the Use of Different Standards for Evaluating Comprehension

Lexical Standards of Evaluation

Applying a lexical standard involves evaluating one's understanding of individual words. The context of these words has an impact on a reader's understanding of the word itself and of the passage as a whole. Comprehension monitoring research has shown that younger and poorer readers over-rely on lexical standards, deciding that they do not understand a text if they do not understand a specific word (Baker, 1985). Analogous neuroimaging studies of texts with known versus unknown vocabulary have not been conducted. The most relevant line of neuroimaging research concerns lexical ambiguity.

In one such lexical ambiguity study, Mason and Just (2007) compared neural responses to balanced homophones (i.e., a word with two equally likely meanings) and biased homophones (i.e., a word with two possible meanings, one dominant and one subordinate). For example, the word *ball* is biased. The dominant meaning of ball refers to sports, but the subordinate meaning, as in a dance, is the intended meaning in the sentence *"This time the ball was moved because it was always so well attended."* When participants read ambiguous words of either type presented in a sentence context, activation was greater in the left inferior frontal gyrus (IFG) than when reading non-ambiguous words. When the ambiguous word was balanced, only the left IFG was activated, but with a biased ambiguous word, the right IFG was also activated. Balanced words result in the generation of multiple meanings, one of which is confirmed as the sentence continues. When readers encounter a biased ambiguous word, they usually select the dominant meaning, and then need to reprocess the text when the subordinate meaning subsequently appears to be correct (a *garden path* effect). Retrieval of the alternate meaning activates the left IFG above and beyond the activation created by a balanced word. In this study, as well as in those discussed throughout the article, participants were young adults, typically college students.

In summary, the lexical standard comes into play when ambiguous words are encountered. Such words need additional neural processing for generation, maintenance and selection of multiple meanings. As Mason and Just (2007) suggested, the additional activation in the IFG for biased ambiguous words may signal increased coherence monitoring (comprehension monitoring) to make sense of the sentence.

Syntactic Standards of Evaluation

Applying syntactic standards involves evaluating one's understanding with respect to the grammaticality of a sentence or the syntactic acceptability of particular words within context. Comprehension monitoring research has shown age and ability related differences in readers' detection of syntactic violations (Baker, 1985). Neuroimaging research has revealed which regions of the brain are active when syntactic challenges are encountered. In this section we consider illustrative fMRI studies of syntactic ambiguity (Fiebach, Vos, & Friederici, 2004; Mason, Just, Keller, & Carpenter, 2003; Rodd, Longe, Randall, & Tyler, 2010).

Syntactic processing difficulties vary according to where in the sentence an ambiguity is located. Mason et al. (2003) had participants read two types of ambiguous sentences, namely main verb ambiguity and reduced relative ambiguity. In the sentence: *"The experienced soldiers warned about the dangers before the midnight raid,"* the verb *"warned"* is the main verb (MV). However, in the sentence *"The experienced soldiers warned about the dangers conducted the midnight raid,"* the verb *"warned"* is in a reduced relative clause (RR). Preliminary work showed that MV parsing is preferred over RR. Participants had more brain activity when reading both types of ambiguous sentences relative to unambiguous ones, with RR constructions eliciting the most activity. These increased activations were in Broca's area, the left inferior gyrus (IFG), and the left superior temporal gyrus (STG).

The ease of processing syntactic ambiguity also depends on where within a sentence it is disambiguated (early or late), as Fiebach et al. (2004) demonstrated. fMRI analyses indicated greater activation for late disambiguation in a large, left-dominant temporo-frontal cortical network and in subcortical regions (thalamus and basal ganglia). Greater activation was also elicited in the right-hemisphere homologue to Broca's area and in the parietal lobe, which suggests that the processing demands of the ambiguous sentences called for additional brain involvement. Activation was also elicited in the STG and superior temporal sulcus (STS), two regions known for their involvement in sentence comprehension, and in the basal ganglia, a region involved in syntactic processing

People are often confronted with the need to evaluate their comprehension on multiple levels simultaneously. Rodd et al. (2010) examined processing of both lexical and syntactic ambiguities. In the lexical condition, ambiguous words were homonyms or homophones. In the syntactic condition, sentences were ambiguous in that they had two syntactic interpretations. Consider the following ambiguity: *"Experience had taught him that boring colleagues damages work relations."* The ambiguity arises because it is unclear whether it is *boring* or *colleagues* that serves as the subject of the sentence. A comparison sentence with only one possible syntactic structure is: *"Experience had taught him that neglecting risks irritates safety experts."* Results revealed activation in the left inferior frontal gyrus (IFG) during both lexical and syntactic ambiguity processing. There was overlap in regions of greatest activity for the two types of ambiguity, with activation found in most of the pars opercularis (Broca's area), the inferior portion of the pars triangularis, and the most posterior portion of the pars orbitalis. Activation was also greater with syntactic difficulty in the left middle temporal gyrus (MTG), including the portion that extends into the angular gyrus (AG) and surpramarginal gyrus (SMG).

In summary, all three studies revealed greater activity in Broca's area with increased syntactic processing demands, consistent with many previous sentence-level studies. On the basis of this work, we may conclude that use of the syntactic standard of evaluation activates Broca's area. Other regions that were activated by syntactic ambiguities were a parietal region, the left IFG, and the left STG. Moreover, the Rodd et al. (2010) study provided evidence that the use of lexical and syntactic standards activates similar regions in the prefrontal cortex.

External Consistency Standards of Evaluation

When readers evaluate their comprehension for external consistency, they are monitoring for consistency with what they already know about the world. Research on comprehension monitoring has examined use of this standard with an error detection paradigm where text is manipulated to contain violations of real world knowledge. External consistency is a standard that is within the repertoire of children under five, but even adults fail to apply it under certain circumstances (Baker, 1989; 2008). Several recent neuroimaging studies have tapped the use of this standard by introducing semantic anomalies into passages. In this section we discuss two types of external consistency studies, those involving "semantic illusions" and those involving processing of counterfactuals.

In Eriksen and Mattson's (1981) seminal research on the so-called "Moses illusion," participants were asked questions such as *"How many animals of each kind did Moses take with him on the ark?"* A majority of people responded "two" even though they had the knowledge that it was Noah, another Biblical figure, who took the animals onto the ark. It seems that participants responded without a full evaluation of what they had heard; in other words, their semantic processing was shallow. Participants did not make errors when names without strong Biblical connections were substituted. Using ERP, Sanford and colleagues (Bohan, Leuthold, Hijikata, & Sanford, 2012; Sanford, Leuthold, Bohan, & Stanford, 2011) tested the possibility that anomalies of the "Moses" type might not yield the N400 response that is characteristic of semantic anomalies because this type of anomaly fits well with the global context. The studies were similar in design, but Sanford et al. used a listening task and Bohan et al. used a reading task. Passages were developed that contained anomalies that were either a good fit to context or a poor fit. An example of a passage with a good fit anomaly is:

Child abuse cases are being reported much more frequently these days. In a recent trial, a 10 year sentence was given to the victim, but this was subsequently appealed.

It is anomalous for a *sentence* to be given to a victim, but not for the victim to receive a *care order*, the term that was used in the consistent version of the passage. An example of a passage with an anomaly that is a poor fit to context follows:

Leon was the manager of a struggling record shop. Yesterday, the owner told him that he would have to think of new ways to sell more letters if he wanted to keep his job.

Selling letters in a record shop is anomalous; a non-anomalous version of the passage replaces *letters* with *records.* In both studies, participants detected almost all of the poor fit anomalies (94%), but only about two thirds of the good fit anomalies were detected.

Participants in the Sanford et al. (2011) study listened to the sentences while ERPs were being measured, and they indicated with a response box whether or not they detected an anomaly. For the good fit anomalies, results revealed larger late positivity potential P600 waveforms for detected anomalous words as compared to non-detected anomalous words and compared to standard words. Consistent with predictions, no N400 effect was found for the good fit anomalies. However, the poor fit anomalous words did elicit an N400 effect relative to the corresponding non-anomalous words. In the Bohan et al. (2012) study, the first sentence, which established the context, was presented on the screen as a whole, and the second sentence was presented word by word. For poor fit anomalies compared with non-anomalous controls, the classic N400 effect was again obtained. In contrast to the listening study, however, both detected *and* non-detected good fit anomalies elicited an N400 effect compared to non-anomalous controls. Thus, it appears that participants processed the anomaly at some level without consciously detecting it. Similar results have come from eye-tracking studies; Daneman, Lennertz, and Hannon (2007) found that readers had longer fixation times on semantic illusions without reporting them as such.

The way readers process counterfactual sentences also reveals how external consistency impacts understanding. In two ERP studies by Nieuwland (Nieuwland, 2013; Nieuwland & Martin, 2012), the focus was on whether prior knowledge of true events would interfere with comprehension of counterfactuals. Nieuwland and Martin used materials based on commonly known historic events, presented as "real-world" or "counterfactual" sentences that were either true or false. For example, a counterfactual true sentence is: *"If N.A.S.A. had not developed its Apollo Project, the first country to land on the moon would have been Russia surely"* and a counterfactual false sentence is: *"If N.A.S.A. had not developed its Apollo Project, the first country to land on the moon would have been America surely."* The first clause of the sentence was presented in its entirety and the second clause was presented word by word on the screen. There was a larger N400 waveform when participants read counterfactual false sentences and real-world false sentences compared to counterfactual true and real-world true sentences, with the magnitude of the effect similar for the two types of false sentences. This indicates that knowledge of real-world events did not make processing counterfactual false sentences more difficult. The interpretation depended on the context of the sentence being read (i.e., counterfactual or real-world), not on previous factual knowledge. Thus, readers made decisions regarding the truth-value of the propositions within the immediate context.

In the follow-up study, Nieuwland (2013) first had participants read sentences that were anomalous with respect to real-world knowledge (e.g., *"Dobermans would breathe under water"*) and confirmed that they elicited the expected N400 effect. A second group of participants then read these sentences preceded by a counterfactual conditional premise that rendered the 'locally anomalous' sentence true (e.g., *"If dogs had gills, Dobermans would breathe under water"*). Results mirrored those of Nieuwland and Martin (2012), where counterfactual-false sentences and real-world-false sentences elicited similar N400 responses compared to counterfactual and real-world true sentences. This suggests that if sufficient and plausible context is added to propositions with unusual meaning, comprehension of the entire sentence poses no difficulties.

In summary, the neuroimaging studies reveal neural correlates of monitoring comprehension for external consistency. An N400 effect was detected in all of the studies in response to violations of real world knowledge, but there were inconsistencies across two of the studies related to the type of semantic anomaly. Poor fit anomalies consistently elicited the N400 effect. In Sanford et al., good fit anomalies (i.e., those that evoke the "Moses illusion") elicited a P600 effect but not an N400 effect. In Bohan et al., however, good fit anomalies elicited a N400 effect, even when they went undetected by the participants. The reason for this discrepancy is unclear, but it may be that the monitoring processes are different for reading versus listening. None of the studies covered in this section used fMRI, but a study to be considered subsequently included semantic anomalies as well as referential ambiguities (Nieuwland, Petersson, & Van Berkum, 2007). The semantic anomalies elicited increased activation in the left and right inferior gyrus (IFG), indicative of the increased semantic processing needed to attempt to interpret the sentences.

Internal Consistency Standards of Evaluation

Monitoring comprehension using an internal consistency standard is a critical component of establishing a coherent representation of a text. A reader or listener using this standard is likely to notice if ideas in a text are inconsistent with one another. Many comprehension monitoring researchers have used an error detection paradigm where internal inconsistencies (or other types of problems) are embedded within a text. Sometimes the inconsistencies are very blatant and easily detectable, but other times they are more subtle. The more subtle violations may be resolvable by making inferences. In this section we consider an fMRI study that used the error detection paradigm (Ferstl, Rinck, & von Cramon, 2005), and we also examine a related line of research on inferential processing (Kuperberg, Lakshmanan, Caplan, & Holcomb, 2006; Virtue, Haberman, Clancy, Parrish, & Beeman, 2006). Readers monitoring their understanding will be aware of the need to make an inference to integrate seemingly inconsistent or incompatible propositions.

The Ferstl et al. (2005) study is described in substantial detail because it provides the most explicit neuroimaging evidence of comprehension monitoring. The study was situated within the discourse processing tradition, but it examined the two component processes of comprehension monitoring, evaluation and regulation. Materials consisted of short stories with inconsistencies involving either the timing of plot events or the emotional state of a character. Participants listened to the stories in the fMRI scanner and made judgments as to whether or not they contained inconsistencies. The following is an example of a story containing a chronological inconsistency:

> *Today, Markus and Claudia would finally meet again. Markus's train arrived at the station 20 minutes before Claudia's train. Markus was very excited when his train stopped at the station on time. He tried to think of what he should say when he met her. Many people were crowding on the platform. Claudia was already waiting for him when he got off the train with his huge bag. They were both very happy.*

In the consistent version, the second sentence stated that Markus's train arrived *after* Claudia's, rather than *before.* The critical content, which was the same in both story versions, was presented in the sixth sentence, *already waiting.* Contrast the chronological type of story with the following, which contains an emotional inconsistency:

> *The semester was finally over and Sarah wanted to celebrate. A lot of her friends had shown up for her end-of-school party. It was one of these parties with everything being just perfect. Sarah's best friend gave her a hug and told her how much fun she was having. Sarah couldn't remember that she had ever been so sad before. She put her favorite record into the CD player and started dancing by herself.*

The consistent version of this passage had the word *happy* instead of *sad* in the fifth sentence. In this case, the inconsistency does not involve two sentences with conflicting information, as in the chronological example, but rather a specific emotional term that is inconsistent with the overall content of the story. Moreover, unlike the chronological inconsistencies, the emotional inconsistencies are potentially resolvable. Listeners or readers could draw on prior knowledge to make a gap-filling inference (e.g., perhaps just after her best friend hugged her, Sarah's boyfriend arrived and broke up with her). These differences led Ferstl et al. (2005) to expect that integration in the chronological stories would involve re-accessing the temporal information presented in the earlier sentence and comparing it with the new information; this process likely would involve prefrontal activation. To create an integrated representation of the emotional stories, areas of the brain engaged during inferential processing would likely be activated, e.g., the right frontotemporal region and/or or the dorsomedial prefrontal cortex (dmPFC).

In order to separate the processes of detection and integration (evaluation and regulation, in comprehension monitoring terms), Ferstl and colleagues (2005) analyzed the images over two different time periods. For detection they examined activations at the point when participants were processing the consistent or inconsistent target information, and for integration they examined activations from the point of target presentation through the end of the story. At the detection point, the right anterior temporal region (aTL) showed increased activation when inconsistencies were presented relative to consistent passages. The involvement of the aTL in language comprehension has been well-established, and the researchers interpreted the right hemisphere involvement as supporting left-hemisphere language areas when processing difficulties arise. As expected, inconsistencies in the two story types elicited activity in different regions of the brain. For the chronological stories, integrating an inconsistency required the contribution of the lateral prefrontal areas in both hemispheres. In contrast, integrating inconsistent emotional information activated the dmPFC, the upper portion of the right IFG, and the precentral sulcus.

In support of the integration (regulation) process, analyses of activation during the period from target word to end of story showed that the dmPFC was highly activated during the last few seconds of each story type. However, activation leveled off last for the inconsistent emotional stories. This was interpreted to mean that when an emotional inconsistency was present that could potentially be resolved, readers engaged in inferential processing. The ventrolateral prefrontal cortex (vlPFC) was activated bilaterally for inconsistent versions of both types of stories. Ferstl et al. (2005) noted that the vlPFC is sensitive to violations of expectations. Activation in this region was somewhat stronger for the chronological stories, suggesting that a reinstatement search was undertaken to connect the two conflicting statements.

An important additional contribution of this study to the comprehension monitoring literature is the contrast of listeners who detected more of the inconsistencies with those who detected fewer. Participants who detected fewer inconsistencies had greater engagement of the dmPFC when reading inconsistent passages than when reading consistent passages. This suggests that individuals who are less certain of their understanding engage in additional processing. Ferstl et al. (2005) cited evidence that dmPFC activation is not caused directly by the external stimulus but rather by internal evaluation processes. These same evaluation processes may be present when readers who do not report a particular inconsistency nonetheless spend more time reading the words of text that set it up (Baker, 1989).

We turn now to consideration of fMRI studies that focus on inferential processes more generally. Virtue et al. (2006) had participants listen to narrative passages with episodes that either did or did not require an inference for successful comprehension. Consider the following excerpt:

> *After the ceremony, John and Nancy stood outside, their hands full of rice, and waited to see the newlyweds. Soon the rice was in their hair and clothes and John took pictures, while Nancy wished her friend a happy honeymoon.*

In this implicit excerpt, one needs to infer that John and Nancy threw the rice at the newlyweds when they came out. Without such an inference, there would be a coherence break after the clause, *"soon the rice was in their hair."* The explicit version of this episode contained the phrase, *"threw it at the newlyweds"* rather than *"waited to see the newlyweds."* Of particular interest in the analysis of the fMRI images was the contrast in activation between the two different passage versions at the point of the coherence break. Results revealed increased fMRI signal for implied over explicit events within the left superior temporal gyrus (STG). Areas within this region are important for the semantic integration of inferential information.

In an fMRI study that involved reading rather than listening, Kuperberg et al. (2006) compared processing of passages that varied in coherence. The first two sentences of each passage provided context and the third sentence was either highly related, intermediately related, or unrelated to the other sentences. The highly related scenario was explicit in causality, as in:

> *The boys were having an argument. They began hitting each other. The next day they had bruises.*

The intermediately related scenario required the reader to make an inference. In the following example, the reader must infer that the boys hit each other:

> *The boys were having an argument. They became more and more angry. The next day they had bruises*.

The unrelated scenario contained no clear inference for the reader to make and thus the third sentence could not be integrated with the other two, as in:

> *The boys were unsure about the weather. At noon they started to hike. The next day they had bruises.*

In the scanner, the first two sentences were presented sequentially, and the third sentence was presented word by word. Participants used a response box to indicate the degree of relatedness of the sentences. As expected, participants required more time to make judgments about the intermediately-related scenarios. In addition, these scenarios elicited increased activity in the right inferior prefrontal gyrus (IFG) and the superior/medial prefrontal gyrus (SMFG), as well as increased activity in the left lateral temporal inferior parietal/prefrontal network (aTL). The activation of the posterior parts of the IFG may reflect additional working memory demands. Increased activation in the SMFG is associated with establishing coherence, suggesting that participants did make the necessary inference.

In summary, the research reviewed here provides compelling evidence that readers and listeners evaluate text for internal consistency. Distinct patterns of neural activity are elicited when passages contain information that is inconsistent with other portions of the passage or that cannot be readily integrated. Explicit inconsistencies activated the language network in the aTL and various regions of the prefrontal cortex, including the dmPFC and the vlPFC. Moreover, Ferstl et al. (2005) showed that different regions of the brain were involved in the detection (evaluation) and integration (regulation) components of comprehension monitoring, consistent with Fernandez-Duque's analysis of the role of executive control regions in metacognition. Prefrontal regions were also activated when inferences were required, including the IFG and the SMFG.

Propositional Cohesiveness Standards of Evaluation

When readers evaluate their understanding with respect to propositional cohesiveness, they are evaluating their ability to form an integrated representation of the local text (Baker, 1985). Failure to successfully link propositions often results from vague or ambiguous anaphoric referents. In this section we discuss two of many neuroimaging studies that have revealed effects of anaphora on text processing (Nieuwland et al., 2007; Nieuwland & Van Berkum, 2008). These two were selected because they contrasted anaphora processing with processing of semantically anomalous text, and as such required readers to use two different semantic standards of evaluation.

Nieuwland and Van Berkum (2008) measured ERPs as participants read stories, presented word by word, in which the target noun phrase was semantically coherent, referentially ambiguous, semantically incoherent, or both incoherent and ambiguous. Consider first the referentially ambiguous version of one of the passages.

> *Police officer Laura had arrested two voyeurs in the Vondelpark, a pushy one and another one that was sneaky. Laura brought both men to the police headquarters. She arranged a cell for the voyeur so the park would remain safe.*

One does not know which voyeur was put in jail. Following is the referentially clear version:

> *Police officer Laura had arrested a pushy thief and a sneaky voyeur in the Vondelpark. Laura brought both men to the police headquarters. She arranged a cell for the voyeur so the park would remain safe.*

The semantically incoherent version follows:

> *Police officer Laura had arrested a pushy thief and a sneaky voyeur in the Vondelpark. Laura brought both men to the police headquarters. She locked them up in the voyeur so the park would remain safe.*

One does not lock criminals up in voyeurs. The fourth version of this set was both semantically incoherent and referentially ambiguous:

> *Police officer Laura had arrested two voyeurs in the Vondelpark, a pushy one and another one that was sneaky. Laura brought both men to the police headquarters. She locked them up in the voyeur so the park would remain safe.*

Results revealed that the ambiguous anaphor produced a frontal negative ERP effect, Nref, characteristic of referential ambiguity, and that semantic incoherence produced a classic N400 effect. The response to the target noun phrase in sentences that were both ambiguous and incoherent resembled that of semantic incoherence alone. Nieuwland and Van Berkum (2008) suggested that semantic incoherence may hinder readers' ability to consider the two competing interpretations produced by the ambiguous reference. Readers who come across a semantic anomaly may spend their time trying to resolve the problem and may gloss over the less evident anaphoric ambiguity. A late positive posterior ERP shift was evident for all but the clear passage versions; the authors suggested this was a P300 response reflecting the extent to which participants were monitoring implausibility or unacceptability of the presented situations in the real world. In other words, this appears to reflect comprehension monitoring.

A similar study examining processing of both anaphora and semantic anomaly was conducted in the fMRI scanner by the same research team (Nieuwland et al., 2007). Participants were presented with sentences of four types: referential ambiguity, *"Ronald told Frank he had a positive attitude towards life"*, referential failure, *"Rose told Emily that he had a positive attitude towards life"*, referential coherence and semantic coherence, *"Ronald told Emily that he had a positive attitude towards life"*, and referential coherence and semantic anomaly, *"Ronald told Emily that he had a positive potato towards life."* The three problematic sentence types were compared to the clear and coherent type. Referential ambiguity elicited increased activation in the medial and bilateral parietal, medial frontal and right superior frontal regions, which Nieuwland et al. (2007) interpreted as evidence that readers engaged in problem solving to comprehend the sentence. Referential ambiguity also elicited activation increases in the dorsal frontal cortex (dMFC) and the anterior/ventral medial frontal cortex (vmPFC), which are involved in making controlled higher order inferences and evaluative judgments. This effect is analogous to the frontal Nref shift that occurs in ERP studies as participants attempt to resolve the ambiguous reference. Referential failure elicited increases in medial and bilateral parietal and left middle frontal brain regions, suggesting to the authors that readers recognized and accepted that referential failure was due to a syntactic error. Semantic anomalies elicited increased activation in the left and right inferior gyrus (IFG), indicative of the increased semantic processing needed to comprehend the sentence. These results provide clear evidence that readers use different neural mechanisms in response to different types of comprehension difficulties.

In summary, studies of referential ambiguity show that readers are sensitive to disruptions in propositional cohesiveness. ERP research revealed Nref effects in response to referential ambiguity (Nieuwland & Van Berkum, 2008), and fMRI research revealed that the medial and bilateral parietal and left medial frontal regions were activated by referential ambiguity and referential failure (Nieuwland et al., 2007).

Neuroimaging Evidence of Individual Differences in Comprehension Monitoring

Individual differences in comprehension monitoring skills have been well documented in behavioral research. The consistent correlational evidence of better performance by more skilled readers has led to the development of interventions to enhance metacognitive knowledge and control of less skilled readers. The success of these interventions in improving reading comprehension has, in turn, led to international mandates for metacognitively-oriented classroom reading instruction (Baker, 2011). Despite the interest in individual differences among educational and developmental psychologists, relatively few neuroimaging studies have examined individual differences in neural responses to challenging text. A few of the studies presented above have done so, and we briefly consider them here.

Two studies revealed differences related to working memory, one addressing syntactic comprehension challenges and the other semantic. Working memory was assessed using a reading span task (Daneman & Carpenter, 1980), which requires readers to hold information in memory while continuing to process text. Fiebach et al. (2004) found that participants with weaker working memory had relatively greater activation in Broca's area, which indicates more effort was required for syntactic processing. Virtue et al. (2006) found that participants with a high working memory span showed greater activation in the left inferior frontal gyrus (IFG) for implied events than for explicit events at coherence breaks in the stories. Although greater activation in the IFG may be taken as evidence of greater effort, Virtue et al. interpreted the activation to reflect the selection and generation of appropriate inferences.

Neuroimaging studies that take into account individual differences in reading skill are even more limited. We discuss one such study not presented earlier. Prat, Mason, and Just (2011) first assessed participants' reading vocabulary. Subsequently, participants read short passages with two sentences that were either moderately or distantly causally related. For example *"Sandra walked barefoot on the littered beach. She had to clean out the wound on her foot"* is a moderately-related passage, whereas *"Tom decided to run a marathon for charity. He had many visitors at the hospital"* is distantly related. Less skilled readers showed greater activation in the left posterior temporal/parietal regions and the right hemisphere homologues of the left hemisphere language regions. The so-called "spillover hypothesis" suggests that language functions in the right hemisphere are less efficient counterparts to language functions in the left hemisphere. The left hemisphere is activated by language first, and the right hemisphere serves as a resource reserve. When the processing demands of a particular language task become too great for the left hemisphere, the residual processing spills over, activating the right hemisphere. The increased right hemisphere activation observed in less-skilled readers suggests that they may be less efficient in their comprehension processes and may fail to make optional inferences.

Neuroimaging Evidence of Comprehension Monitoring while Using Reading Strategies

It is well established that comprehension can be improved when readers use strategies that include metacognitive components (Baker & Beall, 2008; McNamara, O'Reilly, Rowe, Boonthum, & Levinstein, 2007). Two related studies by Moss and colleagues (Moss, Schunn, Schneider, & McNamara, 2013; Moss, Schunn, Schneider, McNamara, & VanLehn, 2011) demonstrate that the use of such strategies elicits different neural activation patterns than the use of strategies involving less cognitive effort. Participants read expository text in the scanner and used three different types of strategies to help learn the material. The simplest strategy, known to be of limited effectiveness, was rereading. A second strategy, which requires some cognitive effort, was paraphrasing the text. The most complex strategy, known to be quite effective in enhancing reading comprehension, was self-explanation (McNamara et al., 2007). In self-explanation reading training, five strategies are taught to students: comprehension monitoring, paraphrasing, elaboration, bridging, and prediction. Participants received training via iSTART, an intelligent tutoring system (McNamara et al., 2007), on using self-explanation prior to the scanning session. In the 2011 study, participants read biology texts, and in the 2013 study they read physics texts with diagrams. After reading a section of each text, participants orally performed the assigned reading strategy, before going on to a section of the next text and performing a second strategy, followed by the third. This process was repeated three times for each strategy.

The taped recordings of the strategies were analyzed to verify that participants used the strategies they were instructed to use. If the self-explanation protocols did not reveal any strategies other than paraphrasing, the session was reclassified as paraphrasing. Neither article reported data on the breakdown of strategy use in self-explanation. The elaboration and prediction components of self-explanation are often included in metacognitive reading strategy training packages along with comprehension monitoring (e.g., reciprocal teaching includes clarifying, activating prior knowledge, and prediction).

Of particular interest in the imaging data were regions of the brain involved in executive control, associated with the application of the strategies, and regions involved in higher level comprehension processes, such as cohesion building. Three planned comparisons were conducted: rereading vs. self-explanation; rereading vs. paraphrasing; and self-explanation vs. paraphrasing. Consistent with hypotheses, the two strategies requiring more cognitive effort were associated with greater activation of the control network than rereading, but the two strategy conditions did not differ. This lack of difference was surprising, given the expectation that a more metacognitively-oriented approach would put greater demands on executive control processes. The self-explanation condition elicited more activation in the cohesion building regions than the paraphrasing condition, however, including the posterior cingulate cortex and the angular gyrus. This was expected because self-explanation explicitly calls for readers to make elaborative and bridging inferences.

The Moss et al. (2013) study replicated results of the Moss et al. (2011) study. It also included an analysis of mind wandering, known to activate what is known as the default network, a frontal area that overlaps with the executive control network. Participants rated the degree to which their mind wandered after completing each strategy session. More mind wandering was reported in the rereading condition, where cognitive demands are lower, than in the other two conditions. Awareness of mind wandering can trigger comprehension repair, which corresponds to the regulation component of comprehension monitoring (e.g., the experience of realizing one has not been paying attention to the text that was just "read" elicits more focused reading).

In summary, readings strategies that require cognitive effort activated regions associated with executive control, but the complexity of the strategies was not associated with differential activation. This lack of a complexity effect may indicate that participants did not invoke comprehension monitoring while using self-explanation. Self-explanation strategies activated the posterior cingulate and the angular gyrus, consistent with the demands of that strategic approach to create bridging and elaborative connections. Greater mind-wandering occurred during rereading, reflecting the more limited cognitive demands of that strategy.

Summary, Conclusions, and Future Directions

In this article we presented a selective synthesis of recent neuroimaging research re-interpreted from a metacognitive perspective to show that adult readers and listeners apply several different standards or criteria for monitoring their comprehension. Most of the research focused on use of a particular standard, whether at the lexical, syntactic, or semantic level (including external consistency, internal consistency, and propositional cohesiveness). What is clear from these studies is that different kinds of challenges to comprehension elicit different patterns of brain activity, whether reflected in electrical activity measured through ERP or hemodynamic activity (blood flow) measured through fMRI. Moreover, the two components of comprehension monitoring, evaluation and regulation, were shown to be separable, eliciting neural activity in different regions of the brain involved in executive control.

Regions of greater neural activation for lexical and syntactic ambiguities, for violations of world knowledge, for inconsistent text, and for referential ambiguities were consistent with those identified in previous fMRI research. Both the extended language network and the executive control network were involved in problem detection and/or correction, regardless of problem type. The inferior frontal gyrus

(IFG), for example, was activated in response to all text manipulations that reduced comprehensibility. The ERP waveforms recorded in response to problematic text were also consistent with previous research, such as the classic N400 effect for semantic anomalies and the Nref for referential ambiguities. It is clear that neuroimaging research complements and extends the behavioral research on the cognitive and metacognitive processes that are engaged during comprehension.

The increasing interest in educational neuroscience (Block & Parris, 2008) and the increasing availability of fMRI scanners at major research institutions is likely to foster initiatives aimed at improving academic attainment. Few existing studies address individual differences related to reading skills or executive control skills, but educational researchers are likely to promote further work along these lines. Despite the practical challenges, longitudinal fMRI research on the development of the executive control network in relation to behavioral changes in metacognition would be invaluable. The fMRI research on reading strategies (Moss et al., 2011) leads us to envision an experiment designed to discover whether training in comprehension monitoring and self-regulated learning strategies could increase the efficiency of neural functioning. The potential contributions are great.

References

Ardila, A. (2013). Development of metacognitive and emotional executive functions in children. *Applied Neuropsychology: Child, 2*, 82-87.

Baker, L. (1985). How do we know when we don't understand? Standards for evaluating text comprehension. In D.L. Forrest-Pressley, G.E. MacKinnon, & T.G. Waller (Eds.), *Metacognition, cognition, and human performance* (pp. 155-206). New York: Academic Press.

Baker, L. (1989). Metacognition, comprehension monitoring and the adult reader. *Educational Psychology Review, 1*, 3-38.

Baker, L. (2008). Metacognitive development in reading: Contributors and consequences. In K. Mokhtari & R. Sheorey (Eds.), *Reading strategies of first- and second-language learners: See how they read* (pp. 25-42). Norwood, MA: Christopher Gordon.

Baker, L. (2011). Metacognition. In V.G. Aukrust (Ed.), *Learning and cognition in education* (pp. 128-135). Oxford: Elsevier.

Baker, L., & Beall, L.C. (2008). Metacognitive processes in reading comprehension. In S.E. Israel & G.G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 373-388). New York: Routledge.

Block, C.C., & Parris, S.R. (2008). Using neuroscience to inform reading comprehension instruction. In C.C. Block & S.R. Parris (Eds.), *Comprehension instruction: Research based practices* (pp. 113-126). New York: Guilford Press.

Bohan, J., Leuthold, H., Hijikata, Y., & Sanford, A.J. (2012). The processing of good-fit semantic anomalies: An ERP investigation. *Neuropsychologia, 50*, 3174-3184.

Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.

Daneman, M., Lennertz, T., & Hannon, B. (2007). Shallow semantic processing of text: Evidence from eye movements. *Language and Cognitive Processes, 22*, 85-105.

Eriksen, T.A., & Mattson, M.E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior, 20*, 540-552.

Fernandez-Duque, D., Baird, J.A., & Posner, M.I. (2000). Executive attention and metacognition regulation. *Consciousness and Cognition, 9*, 288-307.

Ferstl, E.C., Neumann, J., Bogler, C., & von Cramon, D.Y. (2008). The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping, 29*, 581-593.

Ferstl, E.C., Rinck, M., & von Cramon, D.Y. (2005). Emotional and temporal aspects of situation model processing during text comprehension: An event-related fMRI study. *Journal of Cognitive Neuroscience, 17*, 724-739.

Fiebach, C.J., Vos, S.H., & Friederici, A.D. (2004). Neural correlates of syntactic ambiguity in sentence comprehension for how and high span readers. *Journal of Cognitive Neuroscience, 16*, 1562-1575.

Hacker, D.J., Dunlosky, J., & Graesser, A.C. (Eds.). (2009). *Handbook of metacognition in education*. New York, NY: Routledge.

Houde, O., Rossi, S., Lubin, A., & Joliot, M. (2010). Mapping numerical processing, reading, and executive functions in the developing brain: An fMRI meta-analysis of 52 studies including 842 children. *Developmental Science, 13*, 876-885.

Klingsberg, T., Vaidya, C.J., Gabrieli, J.D.E., Moseley, M.E., & Hedehus, M. (1999). Myelination and organization of the frontal white matter in children: A diffusion tensor MRI study. *NeuroReport, 10*, 2817-2821.

Kuperberg, G.R., Lakshmanan, B.M., Caplan, D.L., & Holcomb, P.J. (2006). Making sense of discourse: An fMRI study of casual differencing across sentences. *Neuroimage, 33*, 343-361.

Landi, N., Frost, S.J., Mencl, W.E., Sandak, R., & Pugh, K.R. (2013). Neurobiological bases of reading comprehension: Insights from neuroimaging studies of word-level and text-level processing in skilled and impaired readers. *Reading and Writing Quarterly, 29*, 145-167.

Mason, R.A., & Just, M.A. (2007). Lexical ambiguity in sentence comprehension. *Brain Research, 1146*, 115-127.

Mason, R.A., Just, M.A., Keller, T.A., & Carpenter, P.A. (2003). Ambiguity in the brain: What brain imaging reveals about the processing of syntactically ambiguous sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1319-1338.

McNamara, D.S., O'Reilly, T., Rowe, M., Boonthum, C., & Levinstein, I.B. (2007). iSTART: A web-based tutor that teaches self-explanation and metacognitive reading strategies. In D.S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 397-421). Mahwah, NJ: Erlbaum.

Moss, J., Schunn, C.D., Schneider, W., & McNamara, D.S. (2013). The nature of mind wandering during reading varies with the cognitive control demands of the reading strategy. *Brain Research, 1539*, 48-60.

Moss, J., Schunn, C.D., Schneider, W., McNamara, D.S., & VanLehn, K. (2011). The neural correlates of strategic reading comprehension: Cognitive control and discourse comprehension. *Neuroimage, 58*, 675-686.

Nieuwland, M.S. (2013). "If a lion could speak…" Online sensitivity to propositional truth value of unrealistic counterfactual sentences. *Journal of Memory and Language, 68*, 54-67.

Nieuwland, M.S., & Martin, A.E. (2012). If the real world were irrelevant, so to speak: The role of propositional truth-value in counterfactual sentence comprehension. *Cognition, 122*, 102-109.

Nieuwland, M.S., Petersson, K.M., & Van Berkum, J.J.A. (2007). On sense and reference: Examining the functional neuroanatomy of referential processing. *NeuroImage, 37*, 993-1004.

Nieuwland, M.S., & Van Berkum, J.J.A. (2008). The interplay between semantic and referential aspects of anaphoric noun phrase resolution: Evidence from ERPs. *Brain and Language, 106*, 119-131.

Perfetti, C.A., & Bolger, D.J. (2004). The brain might read that way. *Scientific Studies of Reading, 8*, 293-304.

Prat, C.S., Mason, R.A., & Just, M.A. (2011). Individual differences in the neural basis of casual inferencing. *Brain and Language, 116*, 1-13.

Rodd, J.M., Longe, O.A., Randall, B., & Tyler, L.K. (2010). The functional organization of the fronto-temporal language system: Evidence from syntactic and semantic ambiguity. *Neuropsychologia*, *48*, 1324-1335.

Sanford, A.J., Leuthold, H., Bohan, J., & Stanford, A.J.S. (2011). Anomalies at the borderline of awareness: An ERP study. *Journal of Cognitive Neuroscience*, *23*, 514-523.

Shimamura, A.P. (2008). Neurocognitive approach to metacognitive monitoring and control. In J. Dunlosky & R.A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 373-390). New York: Psychology Press.

Virtue, S., Haberman, J., Clancy, Z., Parrish, T., & Beeman, M.J. (2006). Neural activity of inferences during story comprehension. *Brain Research, 1084*, 104-114.

Received: January 17, 2014

PRAZNA PARNA STRANICA