The time course of building discourse coherence in schizophrenia: An ERP investigation

TALI DITMANa AND GINA R. KUPERBERGa,b
aDepartment of Psychology, Tufts University, Medford, Massachusetts, USA
bDepartment of Psychiatry and Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, Massachusetts, USA

Abstract

Impairments in the buildup and use of context may lead to disorders of thought and language in schizophrenia. To test this hypothesis, event-related potentials (ERPs) were measured while patients and healthy controls read sentences that were highly causally related, immediately related, or unrelated to preceding contexts. Although patients were slower than controls, both groups used the discourse context similarly as evidenced by similar reaction time patterns across conditions. Neurally however, different patterns emerged between patients and controls: within the N400 time window, patients failed to modulate their neural responses across conditions. This failure to differentiate between conditions was specifically correlated with positive thought disorder. Results suggest that schizophrenia patients, particularly those with positive thought disorder, fail to make immediate use of discourse context to build up semantic coherence in the brain.

Descriptors: Schizophrenia, Discourse, ERPs, Language, Inferences

An abnormality in the buildup and use of context to construct a gestalt from individual concepts has been proposed as a primary cognitive deficit underlying the symptoms of schizophrenia (Cohen & Servan-Schreiber, 1992). In the domain of language, this impairment may lead to the disorganized speech, or positive thought disorder, that has long been considered a fundamental clinical feature of schizophrenia (American Psychiatric Association, 1994; Bleuler, 1911/1950). Most studies to date have tested this hypothesis at the level of single words or sentences (for reviews, see Kuperberg & Caplan, 2003; Kuperberg, Ditman, Kreher, & Goldberg, in press).

At the single-word level, studies examining the organization and retrieval of individual items, stored within semantic memory, suggest that there is no overall loss of lexicosemantic knowledge in schizophrenia. Rather, the main abnormalities appear to be in how such knowledge is retrieved or accessed. For example, the use of explicit tasks such as semantic fluency (e.g., Allen & Frith, 1983; Bokat & Goldberg, 2003), word association (e.g., Levine, Schild, Kimhi, & Schreiber, 1996; Moran, Mefferd, & Kimble, 1964), recall (Nestor et al., 1998), and categorization (e.g., Gold, Randolph, Carpenter, Goldberg, & Weinberger, 1992) suggests that the storage and/or access to words in schizophrenia is less organized and structured than in healthy individuals. Additionally, implicit semantic paradigms such as semantic priming suggest that patients’ access to words preceded by semantically related primes may be abnormally increased or reduced, depending on whether experimental conditions bias toward automatic or controlled processing (for reviews, see Kuperberg et al., in press; Minzenberg, Ober, & Vinogradov, 2002).

At the level of whole sentences, patients with schizophrenia are relatively insensitive to semantic anomalies during word monitoring paradigms (Kuperberg, McGuire, & David, 1998, 2000), suggesting that they are impaired in building up sentence context during online processing. Such impairments in the buildup and use of context can, in some cases, lead to processing that is inappropriately driven by the meaning of individual words rather than a whole proposition. For example, Titone, Levy, and Holzman (2000) showed that, unlike healthy controls, patients with schizophrenia were unable to use moderately biasing contexts to inhibit the dominant meanings of homonyms and appropriately select their subordinate meanings. In addition, Kuperberg, Kreher, Goff, McGuire, and David (2006) showed that, in comparison with controls, patients with schizophrenia were relatively insensitive to the introduction of verbs that violated the context of the sentence but that were semantically associated with their preceding argument (e.g., “Every morning for breakfast the eggs would eat . . .”). The same patients, however, were sensitive to contextually violated verbs that were not semantically associated to preceding words in the sentence (e.g., “Every morning for
breakfast the boys would plant . . . ”). Although these impairments in the use of context and susceptibility to lexicosemantic effects can occur even in non-thought-disordered (non-TD) schizophrenia patients, they are often more pronounced in positively TD patients (Bazin, Perruchet, Hardy-Bayle, & Feline, 2000; Kuperberg et al., 1998, 2000; Kuperberg, Kreher, et al., 2006).

In contrast to the studies at the level of single words and sentences, there has been relatively little exploration of how schizophrenia patients build up meaning at the level of discourse (more than one sentence), and most of these studies have focused on the use of linguistic devices to establish referential links during speech production (e.g., Docherty, Cohen, Nienow, Dinzeo, & Dangelmaier, 2003; Rochester & Martin, 1979). Building a coherent representation of discourse meaning, however, also requires the establishment of logical and psychological consistency between the events and propositions described in individual sentences. There is some evidence from memory paradigms that schizophrenia patients fail to use such coherence links across sentences to improve recall of individual sentences (Harvey, Earle-Boyer, Weilgus, & Levinson, 1986; Speed, Toner, Shugar, & Di Gasbarro, 1991), and early studies suggested that patients with low premorbid histories failed to extract the “gist” from groups of individually presented sentences (Knight & Sims-Knight, 1979; although see Grove & Andreasen, 1985). Analogous findings using visual picture stories have recently been described by Brune and Bodenstein (2005).

A problem in interpreting these studies, however, is that they do not necessarily tap into the fast, online neural processes that are engaged as discourse unfolds word by word. It is known that healthy individuals are able to use all information in the discourse context to build up global coherence and that this occurs through fast, online processes in the brain (Ditman, Holcomb, & Kuperberg, in press; Federmeier & Kutas, 1999; Van Berkum, Hagoort, & Brown, 1999). Tapping into the time course of such neural processes is critical for understanding how, moment by moment, patients with schizophrenia make sense of the world around them.

The present study examined the neural basis of building up meaning in discourse using event-related potentials (ERPs) that index neural activity with a millisecond temporal resolution. The focus was on the N400—a negative deflection that peaks at approximately 400 ms after word onset and that reflects the ease of semantically integrating that word into its preceding context (e.g., Holcomb, 1993; for a review, see Kutas & Federmeier, 2000). The N400 is sensitive to conceptual relationships between words within semantic memory (e.g., Bentin, McCarthy, & Wood, 1985; Rugg, 1984), the buildup of meaning within sentences (e.g., Kutas & Hillyard, 1980, 1984), and the buildup of meaning across sentences within whole discourse (e.g., Van Berkum et al., 1999).

In schizophrenia, the degree to which the N400 is modulated by semantic relationships and sentence context (the N400 effect) is sometimes abnormally reduced and sometimes normal. At the level of single words in the semantic priming paradigm, the N400 effect is reduced when experimental conditions bias toward controlled processing (e.g., Condry, Steinhauer, Cohen, van Kammem, & Kasparek, 1999), but, under more automatic processing conditions, it can be normal (Mathalon, Faustman, & Ford, 2002) or, in TD patients, even increased (Kreher, Holcomb, Goff, & Kuperberg, 2007).

At the level of whole sentences, under most circumstances, the N400 is modulated normally in schizophrenia (e.g., Koyama et al., 1994; Kuperberg, Sitnikova, Goff, & Holcomb, 2006). However, under some circumstances, it can be modulated abnormally. First, its modulation can be inappropriately driven by the dominant, rather than the subordinate, meaning of a preceding homonym. For example, in sentences such as “The toast was sincere,” Salisbury and colleagues reported a larger N400 amplitude to “sincere” in patients relative to controls, suggesting that patients interpreted “sincere” as anomalous, having failed to correctly integrate its meaning with the subordinate meaning of its preceding contextual homonym, “toast” (Salisbury, O’Donnell, McCarley, Nestor, & Shenton, 2000; Salisbury, Shenton, Nestor, & McCarley, 2002). This failure to override the dominant meaning of a homonym is evident even when the entire preceding context is consistent with its subordinate meaning: Schizophrenia patients showed an abnormally reduced N400 effect to words (e.g., “river”) within incongruous versus congruous contexts that contained semantically related homonyms (e.g., “bridge”), for example, “The guests played bridge because the river . . . ” versus “Diving was forbidden from the bridge because the river . . . ” (Sitnikova, Salisbury, Kuperberg, & Holcomb, 2002).

The second situation in which the N400 within sentences is modulated abnormally in patients relative to controls is when the semantic anomaly falls on sentence-final words (Adams et al., 1993; Mitchell et al., 1991; Ohta, Uchiyama, Matsushima, & Toru, 1999), where there is increased demand for integrating semantic and syntactic information to “wrap up” the meaning of the entire sentence (Guzman & Klin, 2000). Taken together, these findings suggest that patients are able to use some aspects of context (perhaps the lexicosemantic relationships between individual words), but that they have specific difficulty in using global context to build up and integrate whole sentence meaning.

The current study examined, for the first time, how the N400 is modulated as schizophrenia patients combine information across more than one sentence to integrate a critical word into its entire preceding discourse context. Following a classic behavioral paradigm (e.g., Keenan, Baillet, & Brown, 1984; Myers & Duffy, 1990; Myers, Shinjo, & Duffy, 1987), three types of three-sentence discourse scenarios were constructed. The final sentence was internally coherent but varied in its causal relatedness with its two preceding context sentences. Thus, the three experimental conditions crossed discourse and lexicosemantic influences in the following ways: Final sentences in the highly related scenarios were discourse appropriate and contained critical words that were lexicosemantically related to their preceding context, final sentences in the moderately related scenarios required the generation of a causal inference to be discourse appropriate but also contained critical words that were lexicosemantically related to their context, and the unrelated scenarios were neither discourse appropriate nor contained lexicosemantically related critical words (see Table 1). In addition, because, as mentioned above, previous studies have reported an abnormally reduced N400 effect to sentence-final anomalies in patients, the influence of word position was also examined: Half of the critical words were embedded midsentence and the other half appeared as the sentence-final word.

To determine whether patients construct and use discourse-level representations online while controlling for semantic associative effects of individual words, ERPs to critical words within the final sentences of the moderately related and highly related scenarios were compared. It was predicted that, unlike controls, patients would fail to generate and integrate bridging inferences
online and that the amplitude of the N400 evoked by critical words in the intermediately related scenarios would be the same as to critical words in the highly related scenarios.

To examine how participants integrated both discourse-level and lexico-semantic information across the three sentences within the scenarios, ERPs to critical words within the highly related scenarios were compared with ERPs to critical words within the unrelated scenarios. It was predicted that, unlike controls, patients would fail to use both discourse-level context and lexico-semantic associations to establish discourse coherence, leading to a failure to attenuate the N400 to critical words in the highly related scenarios. Moreover, this neural abnormality was predicted to be most marked in patients with the most severe positive thought disorder.

Methods

Construction of Stimuli

Two hundred and forty sets of three-sentence scenarios, each with highly related, intermediately related, and unrelated conditions, were constructed as described in Table 1. The content words within the highly related and intermediately related scenarios were matched on numbers of word repetitions and on semantic similarity values (SSVs) as calculated using a Latent Semantic Analysis (LSA; Landauer & Dumais, 1997; Landauer, Foltz, & Laham, 1998; available on the Internet at http://lsa.colorado.edu).

These three levels of causal relationships were verified in two ratings studies with healthy volunteers who did not participate in the ERP experiment. These norming studies have been described in detail elsewhere (Kuperberg, Lakshmanan, Caplan, & Holcomb, 2006) but are summarized below for clarity.

Pretest 1: Verification of inference generation. To verify that participants did not generate consistent inferences to the highly related or the unrelated scenarios, but that they did generate inferences to the intermediately related scenarios, discourse scenarios were presented in random order to 12 Tufts undergraduate students. After reading each discourse scenario, participants were asked to write a one-sentence response indicating “why” the event described in the final sentence had occurred. If participants were unable to answer the question, they were told to indicate “don’t know.” Inspection of subjects’ answers indicated that, for the highly related scenarios, subjects wrote responses that were very similar to the second sentence for that scenario, that is, they repeated what they had just read. For the intermediately related scenarios, subjects wrote responses that were very similar to the second sentence of the highly related condition for that scenario, even though they had not seen that sentence, that is, they made the expected inference. For the causally unrelated scenarios, subjects either indicated “don’t know,” or, rarely, they wrote responses that were very different from the second sentence of condition 1 for that scenario, that is, they either failed to make an inference or any inferences generated were inconsistent across subjects.

Pretest 2: Ratings of causal relatedness. An additional 12 participants took part in a rating study in order to (a) obtain ratings of how related the final sentence was to the previous two sentences and (b) objectively determine the word on which subjects made their rating decision—termed the “critical word”—and to ensure that it was the same across the three experimental conditions. Participants were asked to provide a rating of 1, 2, or 3 according to how strongly the final sentence of each scenario was causally related to the preceding two sentences, with 1 indicating a strong causal relationship and 3 indicating a weak relationship. In addition, participants were asked to circle one word within the final sentence that indicated whether it was or was not causally related to the preceding context.

As expected, analyses revealed significant differences in subjects’ ratings across the three scenario types, $F(2,18) = 510.9$, $p < .001$. Subjects rated the highly related scenarios ($M = 1.08$,
SD = 0.07) as being significantly more related than the intermediate related scenarios (M = 1.58, SD = 0.27), t(11) = 7.7, p < .001, that were, in turn, rated as significantly more related than the unrelated sentences (M = 2.82, SD = 0.14), t(11) = 23.4, p < .001.

Scenarios were then divided into three counterbalanced lists (each with 240 sentences, 40 in each condition) and then randomized within lists. Every participant encountered each final sentence once only, and, across all participants, the same final sentences were seen in all three conditions. Within each list, half the critical words occurred at the sentence-final position.

Participants

Twenty patients meeting DSM-IV criteria for schizophrenia confirmed using the SCID (Spitzer, Williams, Gibbon, & First, 1992) and chart examination, all receiving stable doses of atypical antipsychotic medication, were initially recruited from the Lindemann Mental Health Center, Boston. Nineteen demographically matched volunteers on no medication and without histories of psychiatric disorders (Spitzer et al., 1992) were initially recruited by advertisement. All participants (controls and patients) were native, primarily monolingual English speakers who had not learned any other language before age 5 years. All patients were right-handed (Oldfield, 1971; White & Ashton, 1976), without histories of head trauma, neurological disorder, substance abuse within 6 months, or histories of substance dependence. Written informed consent was obtained following the guidelines of the Massachusetts General Hospital and Tufts New England Medical Center Human Subjects Research Committees. Clinical assessments were carried out using the Scale for the Assessment of Positive Symptoms (SAPS; Andreasen, 1987), the Scale for the Assessment of Negative Symptoms (SANS; Andreasen, 1989), and the Positive and Negative Syndrome Scale (PANSS; Kay, Fiszbein, & Opler, 1987) as a measure of overall psychopathology. Five patients and 1 healthy control were subsequently excluded from data analysis for not being able to perform the task (see Behavioral Data Analysis).

Demographic and psychopathological data of the remaining 15 patients and 18 controls are summarized in Table 2. Patients and controls were matched closely on gender and race, and there was no significant difference between the groups in age, years of education, or socioeconomic status (SES) as assessed by the Hollingshead Index (p > .10). Controls had a higher premorbid IQ than patients (p < .001) as assessed by the North American Adult Reading Test (A-NART; Blair & Spreen, 1989).

Stimulus Presentation and Task

Each trial began with a 450 ms fixation with a stimulus onset asynchrony (SOA) of 550 ms. The first two sentences were presented successively, each for 3.4 s (SOA: 3.5 s). The third sentence was presented word by word (each word: 450 ms, SOA: 550 ms). The sentence-final word appeared with a period and was followed by a 750-ms blank-screen interval and then a “+”. This cued participants to press one of three buttons (counterbalanced across participants) depending on the relatedness of the final sentence to the previous two sentences, with 1, 2, and 3, respectively, representing highly related, intermediate-related, and unrelated. This delayed response reduced contamination of the ERP waveform by response-sensitive components such as the P300 (Donchin & Coles, 1988) and triggered the onset of the next trial. After a variable number of scenarios (between 9 and 30), the experimenter asked participants a comprehension question about the content of the scenario that they had just read to ensure that they were attending to the meaning of the scenarios. Participants were given 10 practice trials at the start of the experiment.

Electrophysiological Recording

Twenty-nine tin electrodes were held in place on the scalp by an elastic cap (Electro-Cap International, Inc., Eaton, OH); see Figure 1. Electrodes were also placed below the left eye and at the outer canthus of the right eye, to monitor vertical and horizontal eye movements, and on the left and right mastoids. Impedances were kept below 10 kΩ for the eyes and below 5 kΩ at other sites. The EEG signal was amplified by an Isolated Bioelectric Amplifier System Model HandW-32/BA (SA Instrumentation Co., San Diego, CA) with a bandpass of 0.01–40 Hz and was continuously sampled at 200 Hz by an analogue-to-digital converter. The stimuli and behavioral responses were simultaneously monitored by a digitizing computer. ERPs were averaged off-line at each electrode site for each experimental condition using a 100-ms prestimulus baseline and lasting until 1170 ms post-word onset. Trials contaminated with eye artifact (exceeding 50 μV) or amplifier blockage were excluded from analyses. A blink correction program (using principal component analysis) that computed the impact of the blink on the wave forms in each channel (Dale, 1994) was applied to data from 2 patients and 1 control with greater than 40% blinks in a condition. Artifact contamination from eye movement or amplifier blocking led to the rejection of 13.35% (SD = 9.24) of trials for patients and 9.45% of trials for controls (SD = 58.3). Patients and controls did not differ in the number of trials rejected, t(31) = 1.47, p = .15 (see Table 3 for the artifact rejection rates for each condition).

Behavioral Data Analysis

It was first determined that participants were able to perform the task by calculating d′, a measure of participants’ ability to discriminate between the highly related and unrelated conditions. These conditions were chosen for two reasons. First, they were the easiest to discriminate; thus, if a participant was unable to differentiate between them, it would be difficult to interpret their data. Second, for the temporarily related scenarios, responses depended on whether, within the time provided, participants actually generated and integrated the bridging inference required to
make these sentences coherent. Thus, responses to this condition cannot truly be conceptualized as errors but are more subjective (discussed further below). An a priori cutoff of .55 resulted in the elimination of 5 patients and 1 control, leaving 15 patients and 18 controls.

Relatedness judgments were examined by calculating averages for the relatedness judgments for each condition, with trials judged as “highly related” given a score of 1, “somewhat related” trials given a score of 2, and unrelated trials given a score of 3. A 3 (Relatedness: highly related, intermediately related, unrelated) × 2 (Position: mid-sentence critical word, sentence-final critical word) × 2 (Group: controls, patients) mixed model ANOVA, with Relatedness and Position as within-subjects variables and Group as a between-subjects factor, was performed on these data. Planned comparisons were conducted to follow up significant effects.

In addition, reaction times to the probe were examined. For the highly related and unrelated conditions, only response times to correctly answered trials were analyzed. For the intermediately related condition, all trials, regardless of response, were included in analyses. This approach was taken for several reasons: (a) responses to intermediately related scenarios were based on whether participants generated the inference in the time provided; thus, a response of “highly related” to intermediately related scenarios was a subjective judgment and not incorrect per se—rather it suggested that an inference had been generated; (b) for the intermediately related condition, there were relatively few trials in which participants’ responses matched the a priori condition, with four controls and five patients having fewer than 10 trials; (c) because of these behavioral differences, the critical words in the analysis where ERPs were averaged by behavior would not have been counterbalanced across participants across the three conditions, leading to possible confounds in the results due to differences in frequency and number of letters of these critical words. A 3 (Relatedness: highly related, intermediately related, unrelated) × 2 (Position: mid-sentence critical word, sentence-final critical word) × 2 (Group: controls, patients) mixed model ANOVA was performed on these data. Planned comparisons were conducted to follow up significant effects.

**ERP Data Analysis**

All analyses were conducted on the mean amplitudes of ERPs evoked by critical words (using a 100-ms prestimulus baseline) over a 375–500-ms time window. Although a little later than a typical N400, this captured the window in which differences between the conditions were seen in both patient and control groups and was chosen to maximize chances of depicting true group differences. In addition, as a post hoc analysis, modulation within the 700–1000-ms time window, corresponding to the Late Positivity Component (LPC), was also examined to explore observed differences in the waveforms between conditions and groups. Similar to the reaction time analysis described above, for highly related and unrelated conditions, only accurate trials were included in the analyses and for the intermediately related condition, all trials, regardless of response, were analyzed.

To examine front-to-back extent (anterior, central, posterior) of effects, three sites—an anterior, central, and posterior site—in each of the three lateral columns were chosen for analysis; Column 1 included FC1/2, C3/4, P3/4, Column 2 included FC5/6, CP5/6, P3/4, and Column 3 included F7/8, T3/4, T5/6 (see Figure 1). A repeated measures ANOVA conducted on these lateral columns included the between-subjects factor of Group (controls, patients) as well as five within-subjects factors: (1) Column (column 1, column 2, column 3), (2) Anterior-Posterior (AP) Distribution (anterior, central, posterior), (3) Hemisphere (left, right), (4) Relatedness (highly related, intermediately related, unrelated), and (5) Position (mid-sentence critical word, sentence-final critical word). A second ANOVA was conducted on midline sites, which was identical to the analysis at lateral columns with two exceptions: AP Distribution had five levels (FPz, Fz, Cz, Pz, Oz) and there was no Hemisphere factor. A Greenhouse–Geisser correction was applied to all analyses with more than one degree of freedom in the numerator (Greenhouse & Geisser, 1959). In these cases, the original degrees of freedom are reported with the corrected p value.

Interactions with Group were followed up in two ways. First, relative differences between two scenario types were examined in each group separately. Second, the N400 amplitude evoked by a critical word in each scenario type was directly compared between patients and controls.

**Correlations**

Correlations were conducted between (a) the difference in ERP responses evoked at Pz by critical words within the unrelated scenarios and both the highly related and the immediately related scenarios, and (b) total SAPS, SANS, and PANSS scores, positive thought disorder, delusions, and hallucinations as assessed by the SAPS, and chlorpromazine equivalents. Alpha was set to p < .05 for all analyses.

![Figure 1. Electrode montage.](image-url)
Table 3. Artifact Rejection Rates for Patients and Controls

<table>
<thead>
<tr>
<th></th>
<th>Sentence-final</th>
<th></th>
<th>Midsentence</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Controls</td>
<td>Patients</td>
<td>Controls</td>
<td>Patients</td>
</tr>
<tr>
<td>Highly related</td>
<td>10.22 (6.38)</td>
<td>19.60 (11.72)</td>
<td>7.12 (5.62)</td>
<td>10.82 (11.11)</td>
</tr>
<tr>
<td>Intermediately</td>
<td>9.73 (8.44)</td>
<td>17.58 (16.34)</td>
<td>8.03 (6.39)</td>
<td>9.22 (8.66)</td>
</tr>
<tr>
<td>Unrelated</td>
<td>13.40 (13.58)</td>
<td>12.19 (10.93)</td>
<td>8.21 (5.48)</td>
<td>10.67 (11.01)</td>
</tr>
</tbody>
</table>

Notes. Means are given with standard deviations in parentheses.

Results

Behavioral Data

Relatedness judgments (Table 4). An ANOVA revealed a main effect of Relatedness, \( F(2,62) = 451.38, p < .001 \), and a Relatedness \( \times \) Group interaction, \( F(2,62) = 9.02, p < .01 \). Examining controls and patients separately, simple effects ANOVAs demonstrated that both groups behaviorally differentiated between the Relatedness conditions, controls: \( F(2,34) = 640.33, p < .001 \); patients: \( F(2,28) = 95.62, p < .001 \). Follow-up paired \( t \)-tests showed a similar pattern of results across the three conditions for patients and controls: highly related scenarios were judged as being significantly more related than intermediately related scenarios and the unrelated related scenarios were judged as being the least related (all pairwise comparisons in patients: \( ps < .001 \); all pairwise comparisons in controls: \( ps < .001 \)). However, an independent samples \( t \) test comparing responses to each condition separately revealed that patients judged highly related conditions as slightly less related relative to controls, \( t(31) = -2.62, p < .05 \), and as slightly less related relative to controls, \( t(31) = 2.51, p < .05 \). Patients and controls did not differ in judgments to the intermediately related condition, \( t(31) = 0.45, p = .94 \). There were no other significant effects (all Fs < 2.20, all ps > .12).

Reaction times (Table 4). ANOVAs demonstrated main effects of Relatedness, \( F(2,62) = 10.26, p < .001 \). Follow-up \( t \) tests showed that, across both groups, participants took the least time to respond to the highly related scenarios (pairwise comparisons \( ps < .001 \)). RTs to intermediately related and unrelated scenarios did not differ from one another (\( p = .36 \)). Not surprisingly, both participant groups were faster at making their decisions when the critical word appeared midsentence, as reflected by a main effect of Position, \( F(1,31) = 91.02, p < .001 \), although the difference in RTs between midsentence and sentence-final critical words was larger in patients than controls as evidenced by a Group \( \times \) Position interaction, \( F(1,31) = 4.64, p < .05 \). In addition, patients showed overall slower RTs than controls, as reflected by main effects of Group, \( F(1,31) = 13.18, p < .01 \). However, Group did not interact with Relatedness (all Fs < 2.23, all ps > .12), indicating that the patterns of RTs across conditions in the two groups were the same.

ERP Data

\( N400: 375–500 \text{ ms} \). As shown in Figure 2, the N1-P2 complex was followed by a negative-going component (the N400) between 375 and 500 ms in both patient and control groups. Both patient and control groups appeared to show some distinction between the related and unrelated conditions, particularly when the critical word fell in the sentence-final position. A Relatedness \( \times \) Position interaction, midline: \( F(2,62) = 3.16, p < .06 \); lateral: \( F(2,62) = 4.58, p < .05 \), followed up by simple effects ANOVAs demonstrated significant main effects of Relatedness at sentence-final critical words, midline: \( F(2,62) = 16.96, p < .001 \); lateral: \( F(2,62) = 15.20, p < .001 \), and midsentence critical words, midline: \( F(2,62) = 5.47, p < .01 \); lateral: \( F(2,62) = 3.78, p < .05 \).

Of most interest, however, patient and control groups differed in their responses to the relatedness manipulation, as reflected by Relatedness \( \times \) Group interactions, midline: \( F(2,62) = 6.41, p < .01 \); lateral: \( F(2,62) = 7.00, p < .01 \), which remained significant even when excluding trials in the intermediately related condition on which participants indicated that sentences were unrelated, midline: \( F(2,62) = 6.09, p < .01 \); lateral: \( F(2,62) = 5.51, p < .05 \). In the control group, a Column \( \times \) Relatedness \( \times \) Position interaction was observed, lateral: \( F(4,68) = 4.97, p < .05 \). Follow-up simple effects ANOVAs revealed Position \( \times \) Relatedness interactions at all columns except Column 3, midline: \( F(2,34) = 4.43, p < .05 \); Column 1: \( F(2,34) = 5.76, p < .05 \); Column 2: \( F(2,34) = 6.23, p < .02 \); Column 3: \( F(2,34) = 1.86, p = .18 \). Although main effects of Relatedness were observed at these columns for both sentence-final critical words, midline: \( F(2,34) = 23.95, p < .001 \); Column 1: \( F(2,34) = 33.06, p < .001 \); Column 2: \( F(2,34) = 34.86, p < .001 \), and mid-sentence critical words, midline: \( F(2,34) = 18.73, p < .001 \); Column 1: \( F(2,34) = 21.82, p < .001 \); Column 2: \( F(2,34) = 21.13, p < .001 \), a different pattern of results was observed at these positions. At sentence-final critical words, highly related scenarios evoked the smallest amplitude N400, followed by intermediately related scenarios, and unrelated scenarios evoked the largest amplitude N400 (all pairwise comparisons, \( ps < .01 \)). At midsentence critical words, highly related and intermediately related scenarios did not differ from one another.

Table 4. Relatedness Judgments and Reaction Time (RT) Data

<table>
<thead>
<tr>
<th>Relatedness judgments</th>
<th>Sentence-final</th>
<th></th>
<th>Midsentence</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Controls</td>
<td>Patients</td>
<td>Controls</td>
<td>Patients</td>
</tr>
<tr>
<td>Highly related</td>
<td>1.11 (0.10)</td>
<td>1.24 (0.18)</td>
<td>1.12 (0.11)</td>
<td>1.24 (0.16)</td>
</tr>
<tr>
<td>Intermediately related</td>
<td>1.57 (0.29)</td>
<td>1.64 (0.31)</td>
<td>1.53 (0.29)</td>
<td>1.60 (0.25)</td>
</tr>
<tr>
<td>Unrelated</td>
<td>2.84 (0.14)</td>
<td>2.56 (0.45)</td>
<td>2.83 (0.11)</td>
<td>2.55 (0.40)</td>
</tr>
</tbody>
</table>

| RT (in ms)            | Highly related | 1883 (235) | 2612 (735) | 1779 (285) | 2282 (602) |
|                       | Intermediately related | 2422 (435) | 3153 (973) | 2207 (459) | 2724 (681) |
|                       | Unrelated       | 2147 (547) | 2977 (1011) | 1946 (519) | 2912 (1315) |

Notes. Means are given with standard deviations in parentheses.
Building discourse coherence in schizophrenia

Figure 2. ERPs at five electrode sites in healthy adults and schizophrenia patients.

(\(ps > .61\)) and unrelated scenarios evoked the largest amplitude N400 (all pairwise comparisons, \(ps < .001\)).

In contrast, the schizophrenia group failed to show a main effect of Relatedness, midline: \(F(2,28) = 1.62, p = .22\); lateral: \(F(2,28) = .88, p = .41\). Although there was an interaction of Relatedness \(\times\) AP Distribution at lateral columns, \(F(4,56) = 3.71, p < .05\), follow-up ANOVAs did not reveal a main effect of Relatedness at anterior, central, or posterior sites (all \(Fs < 1.99, all\ \(ps > .17\)). Relatedness did not interact with AP Distribution, Hemisphere, and/or Position (all \(Fs < 2.39, all\ \(ps > .11\)).

The second way the Relatedness \(\times\) Group interactions were followed up was by comparing the N400 amplitude between the patient and control groups in each scenario type separately. Figure 3 depicts these differences in highly related, intermediate related, and unrelated scenarios, respectively. In the highly related scenarios, patients evoked more negative amplitude N400s than controls. This difference between controls and patients was most pronounced for sentence-final relative to midsentence critical words, demonstrated by Position \(\times\) Group interactions, midline: \(F(1,31) = 10.30, p < .01\); lateral: \(F(1,31) = 11.54, p < .01\). Follow-up ANOVAs demonstrated main effects of Group for sentence-final critical words (all \(ps < .01\)) and, at lateral columns, midsentence critical words, midline: \(p = .14\); lateral: \(p < .05\). A Group \(\times\) Position \(\times\) Column interaction showed that this effect appeared at all columns except Column 3 (Column 1, Column 2: all \(ps < .01\); Column 3: \(p = .23\)). In addition, it was maximal at centroparietal sites as evidenced by a Group \(\times\) Position \(\times\) AP Distribution interaction, midline: \(F(4,124) = 6.72, p < .01\); lateral: \(F(2,62) = 5.20, p < .05\).

In the intermediate related scenarios, a more negative N400 amplitude in patients at sentence-final words was only evident at posterior sites at the midline column, as determined by a Group \(\times\) Position \(\times\) AP Distribution interaction at the midline column, \(F(4,124) = 4.07, p < .05\). There were no other effects of Group or interactions involving Group in the intermediate related scenarios (all \(Fs < 2.19, all\ \(ps > .10\)). Finally, for the unrelated scenarios, a Column \(\times\) Position \(\times\) AP Distribution \(\times\) Group interaction was observed, lateral: \(F(4,124) = 3.67, p < .05\), with Position \(\times\) AP Distribution \(\times\) Group interactions observed at the midline column, Column 2, and marginally at Column 1, midline: \(F(4,124) = 3.18, p < .05\); Column 1: \(F(2,62) = 2.98, p < .09\); Column 2: \(F(2,62) = 4.66, p < .05\); Column 3: \(F(2,62) = 1.36, p = .26\). Simple effects ANOVAs revealed Position \(\times\) Group interactions at centroparietal sites (FPz: \(p = .24\); FC5/6: \(p < .06\); all other sites \(ps < .05\)) and follow-up independent samples \(t\) tests between controls and patients at these sites demonstrated that patients evoked a more negative N400 amplitude than controls at sentence-final words at parietal sites (Pz: \(p < .05\);

Figure 3. N400 amplitudes at electrode site Pz in schizophrenia patients and healthy adults.
P3/4: $p < .08$; all other sites: $ps > .11$), but not at mid-sentence critical words (all $ps > .18$).

Due to differences in premorbid IQ between patients and controls, analyses were repeated using a smaller sample of 10 controls and 10 patients who were matched on this measure. This analysis again revealed a significant Group × Relatedness interaction, midline: $F(2,36) = 4.16, p < .05$; lateral: $F(2,36) = 3.29, p < .07$. In addition, within the larger sample, premorbid IQ failed to predict the N400 effect to the unrelated (relative to the highly related) critical words (at electrode site Pz, where differences were largest) in either the patient group, $\beta = .099, t(13) = .14, p = .89$, or the control group, $\beta = -.003, t(16) = 0.03, p = .97$, nor did it predict the N400 effect at Pz to the intermediately related (relative to the highly related) critical words in patients, $\beta = .028, t(13) = 0.44, p = .67$, or controls, $\beta = .003, t(16) = 0.03, p = .98$.

**Exploration of the LPC (700–1000 ms).** Consistent with the extant literature, the amplitude of the waveform within this time window was generally more positive on sentence-final than mid-sentence words, reflected by main effects of Position, midline: $F(1,31) = 26.57, p < .001$; lateral: $F(1,31) = 31.35, p < .001$. This effect was evident all over the scalp but was larger at right posterior sites, as reflected by a Position × Hemisphere × AP Distribution interaction, lateral: $F(2,62) = 4.78, p < .05$.

Of most interest, there appeared to be differences between the ERPs to critical words in each scenario type within this time window in patients but not healthy controls. This group difference was reflected by Relatedness × Group interactions, midline: $F(2,62) = 3.91, p < .05$; lateral: $F(2,62) = 3.75, p < .05$, which remained significant even when including 10 controls and 10 patients who were matched on premorbid IQ, midline: $p < .05$; lateral: $p < .05$. In addition, this interaction remained significant even when excluding trials in the intermediately related condition on which participants indicated that sentences were unrelated, midline: $F(2,62) = 3.24, p < .05$; lateral: $F(2,62) = 3.12, p < .06$. For the patient group, follow-up of these interactions revealed marginally significant effects of Relatedness, midline: $F(2,28) = 2.79, p < .09$; lateral: $F(2,28) = 3.25, p < .07$, due to significantly greater positivities to critical words in the unrelated scenarios than in the highly related scenarios ($ps < .05$); no other differences were significant ($ps > .10$). The control group, however, failed to show any differences between the scenario types, midline: $F(2,34) = 1.01, p = .37$; lateral: $F(2,34) = .42, p = .63$.

**Correlations with symptoms and medication dose within the patient group** (Figure 4). Consistent with predictions, higher positive thought disorder scores correlated with smaller N400 amplitude differences between the unrelated and the highly related scenarios at Pz (Spearman $r = -.70, p < .01$) and were also marginally associated with N400 amplitude differences between the unrelated and the intermediately related scenarios (Spearman $r = -.48, p < .08$). There were no other significant correlations with any other symptoms and medication dose (all $ps > .10$).

**Discussion**

The present study examined the neural indices of building up coherence during online discourse comprehension. Schizophrenia patients and healthy adults were presented with three-sentence scenarios with final sentences that were either highly related, intermediately related, or unrelated to their preceding two-sentence contexts. In isolation, the final sentences in all three conditions were coherent. Consistent with previous ERP studies demonstrating that healthy adults are immediately sensitive to both lexico-semantic and discourse-level information as language unfolds online (Ditman et al., in press; Federmeier & Kutas, 1999; Van Berkum et al., 1999), healthy adults showed a decrease in the amplitude of the N400 with an increase in causal relatedness across the three scenario types. The absence of N400 modulation in the schizophrenia group suggests that, between 375 and 500 ms, they failed to use both discourse cues and lexico-semantic information across sentences to build up global coherence. Results remained the same when all intermediately related trials were included in the analysis as well as when trials on which participants indicated that the discourse scenario was unrelated were excluded.

It is unlikely that patients made no attempt to perform the task or to attend to the scenarios. Although patients were slightly less likely than controls to judge highly related scenarios as being “very related” and unrelated scenarios as being “unrelated,” both patients and controls rated highly related scenarios as being most related, unrelated scenarios as being least related, and intermediately related scenarios were judged as intermediately related. In addition, although their behavioral responses were slower than controls, both groups demonstrated a similar pattern of RTs to the relatedness conditions: fastest to the highly related scenarios relative to intermediately related and unrelated conditions, which did not differ from one another. These delayed judgment RTs are likely to have indexed “off-line” neurocognitive processes involved in making the relatedness judgments themselves rather than the online, word-by-word processes indexed by the ERP measures. Thus, they might not necessarily generalize to paradigms in which patients are under pressure to respond (indeed, preliminary findings using a speeded response task showed that, unlike controls, patients take just as long to make relatedness judgments on unrelated scenarios as to intermediately related scenarios; Kuperberg et al., 2005). Nonetheless, in the present study, patients’ normal pattern of behavioral findings suggests that, by the time they were required to make

**Figure 4.** Negative correlation between N400 amplitude differences between unrelated and highly related scenarios associated with positive TD scores. For interpretational ease, the y-axis reflects N400s evoked to unrelated minus highly related scenarios multiplied by −1. Thus, larger values on this axis indicate greater negativity to unrelated scenarios relative to highly related scenarios.
their relatedness decisions, they were attempting to link the final sentences to their preceding contexts.

In addition, the more positive LPC to unrelated relative to highly related critical words provides some evidence that, at a later stage of processing, patients neurally discriminated between the highly related and unrelated scenarios. Interestingly, this LPC was not observed in healthy adults, possibly because they had already attempted to semantically integrate the final sentence into its unrelated context (as reflected by differences in the N400 time window). One explanation for the LPC to the unrelated scenarios in the patients is that it reflected a later inappropriate attempt to draw inferences to the unrelated scenarios. This interpretation would be consistent with recent fMRI findings demonstrating that, unlike controls, schizophrenia patients show an increased engagement of cortical networks to unrelated relative to highly related scenarios (Kuperberg et al., 2005). However, given that the exploration of the LPC was post hoc and that it was not modulated in patients when the analysis was restricted to trials that were judged as consistent with the a priori scenario types, this finding and its interpretation should be viewed as preliminary.

In controls, the N400 amplitudes did not differ between mid-sentence highly related and intermediately related critical words. However, at the sentence-final position, the N400 to intermediately related critical words was more negative than to highly related critical words and is interpreted as reflecting discourse-level, rather than lexicosemantic, integrative processes. This is because the highly related and intermediately related scenario types were matched in terms of lexicosemantic associations between their individual words and, as such, should evoke similar N400 amplitudes if semantic integration was solely modulated by lexicosemantic associations. The discourse-level integrative processes indexed by the N400 to intermediately related critical words are likely to have reflected the increased cognitive effort required to generate and integrate bridging inferences in these scenarios (Keenan et al., 1984; Myers & Duffy, 1990; Myers et al., 1987). For example, to understand the intermediately related scenario, “Mark and John were having an argument. Mark got more and more upset. The next morning John had many bruises,” readers needed to go beyond what was explicitly stated in the text to infer that Mark hit John, resulting in John’s bruises. In addition, the difference in the pattern of activation evoked to mid-sentence relative to sentence-final critical words suggests that the healthy controls did not initially arrive at a discourse-level interpretation but did so by the end of the sentence.

In the schizophrenia group, the absence of an N400 effect in this contrast suggests that, by 400 ms after the onset of critical words, patients had not attempted to generate or use such inferences to build coherence across the three sentences. An alternative explanation for their failure to modulate the N400, however, is that they inappropriately used lexicosemantic associations across sentences to semantically prime and attenuate the N400 evoked by critical words in both the highly related and the intermediately related scenarios. This explanation follows from previous behavioral and electrophysiological studies using semantic priming paradigms that have reported abnormally increased direct (e.g., Moritz et al., 2001; Spitzer et al., 1994) and indirect (e.g., Kreher et al., 2007; Moritz et al., 2001; Spitzer, Braun, Hermle, & Maier, 1993) lexicosemantic priming under automatic experimental conditions, particularly in thought-disordered schizophrenia patients. This account, however, does not fully explain the current findings. If critical words were “hyper-primed” by semantically associated preceding words, this would predict a smaller N400 amplitude to critical words in both the highly related and the intermediately related scenarios in the patients relative to the controls (see discussion by Mathalon et al., 2002). However, the amplitudes of ERPs were larger in patients than controls, suggesting that patients found it generally harder to integrate the meaning of the critical word into its preceding context (for consistent findings at the sentence-level, see Nestor et al., 1997; Niznikiewicz et al., 1997).

In controls, the smaller N400 to critical words in the highly related scenarios, relative to the unrelated scenarios, is interpreted as reflecting the use of both discourse-level and lexicosemantic information to semantically integrate the critical words into the preceding discourse context. In patients, once again, no N400 effect was observed, suggesting that they were relatively insensitive to both types of information. Moreover, in the highly related scenarios, the amplitude of the N400 evoked by critical words in the patients was more negative than in the controls, particularly when the critical word fell at the sentence-final position, perhaps due to the increased processing demands associated with wrapping up and/or evaluating sentences at clause boundaries (Guzman & Klin, 2000). These findings suggest that patients had more difficulty than controls in taking advantage of both congruous lexicosemantic and discourse-level information to semantically integrate the critical word into its discourse context.

As predicted, within the patient group, the severity of positive thought disorder was inversely correlated with the size of the N400 effect to critical words in the unrelated relative to the highly related scenarios. This correlation was specific: There were no other significant correlations with any other measures of psychopathology. This finding is consistent with there being a mechanistic link between these online electrophysiological abnormalities and clinical phenomena such as tangentiality and derailment observed in the language produced by schizophrenia patients. Of note, however, the smaller N400 effect to critical words in the unrelated (relative to the highly related) scenarios in patients relative to controls was not driven entirely by those patients with positive thought disorder, suggesting that some of these neural abnormalities may be associated with schizophrenia as a whole. Clinical thought disorder may reflect a relatively extreme neural abnormality, manifesting only when there is a complete breakdown in the online buildup and use of discourse context (for consistent evidence at the sentence level, see Kuperberg et al., 1998, 2000; Kuperberg, Sitnikova, et al., 2006).

Finally, although patients and controls differed in premorbid IQ (as measured by the A-NART), it seems unlikely that this can explain the absence of N400 effects at the discourse level in the present study. First, as discussed, previous ERP studies at the sentence level have reported normal N400 effects in schizophrenia and, in many of these patient samples, subjects had low average scores on the NART (e.g., Kuperberg, Sitnikova, et al., 2006) and on selective subtests of the WAIS-R (e.g., Nestor et al., 1997). Indeed, normal N400 effects have been observed when patients had significantly lower WAIS-R Information and Vocabulary scores (Sitnikova et al., 2002) and significantly lower NART scores (Kuperberg, Sitnikova, et al., 2006) than healthy controls. Second, results in the present study replicated using a smaller sample size of 10 controls and 10 patients who were matched on A-NART scores. Third, within the patient and control groups, premorbid IQ failed to predict observed neural differences.
An important question raised by these findings is how much a failure to build up discourse coherence online can be attributed to more general cognitive impairments in schizophrenia, such as working memory dysfunction (for a review, see Lee & Park, 2005). In healthy individuals, working memory is known to be engaged during normal inference generation (Klin, 1995, Experiment 3; Singer & Ritchie, 1996) and, at the sentence level, variation in working memory among healthy individuals can account for variability in language function (Caplan & Waters, 1999; Just & Carpenter, 1992) and sensitivity to sentential context as determined by patterns of N400 modulation (e.g., Van Petten, Weckerly, Heather, & Kutas, 1997). Future studies will test the hypothesis that patients’ online neural insensitivity to discourse-level context can be predicted by their working memory capacities, as assessed using more general measures of verbal working memory span (e.g., Caplan & Waters, 1999; Just & Carpenter, 1992).

Conclusions

In sum, the present study demonstrated that schizophrenia patients, particularly those with positive thought disorder, are impaired in immediately using causal information across sentences to build up global representations during online neural processing. Rather, consistent with previous studies (Brune & Bodenstei, 2005; Knight & Sims-Knight, 1979), patients initially focused on local coherence at the expense of the “big picture” global representation. At a later time point, perhaps in response to task demands and strong contextual cues, patients attempted to generate causal inferences in order to differentiate between three levels of causal relatedness, leading to task performance that was strikingly similar to that of healthy adults. Future studies will determine whether these abnormalities are also present in unmedicated patients in their first episode of illness and whether they can be linked to more general cognitive impairments outside the language domain.

REFERENCES


(Received February 9, 2007; Accepted June 1, 2007)