Journal of Abnormal Psychology
1996, Vol. 105, No. 1, 81-88

0021-843X/96/$3.00

Language Production and Thought Disorder in Schizophrenia

Deanna M. Barch and Howard Berenbaum
University of Illinois at Urbana-Champaign

The authors examined the relationship between language production (LP) processes and thought disorder. Thirty-nine schizophrenic or schizoaffective participants completed tasks measuring discourse planning, monitoring, and grammatical–phonological encoding, as well as an interview used to rate thought disorder. The authors found that different LP processes were differentially related to different thought disorder subtypes. Incompetent references were strongly and selectively related to discourse planning performance. In addition, word approximations–neologisms were strongly and specifically associated with grammatical–phonological encoding performance. The article concludes with a discussion of the implications of these results for understanding the multifaceted nature and etiology of thought disorder.

In light of the wealth of research devoted to the study of thought disorder in schizophrenia, progress in elucidating its etiological basis has not been as rapid as one would expect. As characterized by Maher (1991, p. 438), "the study of schizophrenic language has proceeded by fits and starts; hypotheses appear from many different sources, cause a brief stir as they are tested, and then sink from sight as permanently as pebbles thrown into a pond." Despite Maher's pessimistic appraisal, we believe that prior work on thought disorder can potentially lead to major advances in unraveling its etiology. The major impediment to progress in this area may be the lack of comprehensive, yet precise, models of thought disorder. Previous theories may each be able to explain different dimensions of thought disorder. What is missing is a way to integrate across diverse theories.

Several factors have made it difficult to generate comprehensive models of thought disorder. Many models treat thought disorder as a single, global construct, despite evidence to the contrary (e.g., Andreasen, 1982; Berenbaum & Barch, in press; Cutting & Murphy, 1988). For example, various theorists have argued that their models explain "incoherent language output" (Crosson & Hughes, 1987, p. 605), "thought disorder" (Gray, Feldon, Rawlins, Hemsley, & Smith, 1991, p. 2), or "positive thought disorder" (Harrow, Lanin-Kettering, & Miller, 1989, p. 354). In addition, many models focus on a single deficit as the explanation for all aspects of thought disorder. For example, Cohen and Servan-Schreiber (1992, p. 45) argued that a deficit in "the internal representation of context" explains language abnormalities in schizophrenia, and Hoffman (1986, p. 836) suggested that "diminished discourse planning abilities" are responsible for thought disorder. Because thought disorder is not a unitary construct, a single deficit is unlikely to be sufficient to explain all facets of thought disorder. Furthermore, separate dimensions of thought disorder may be related to distinct etiological factors.

We believe that what is needed to advance our understanding of thought disorder is a way to integrate theories that focus on different aspects of language and thought. We believe that the framework of a comprehensive model of language production (LP) can be used to achieve this goal. It seems plausible that thought disorder is related to LP if (a) one defines *thought disorder* as observable language disturbances, making no assumptions about underlying "thought" disturbances; and (b) one defines LP broadly, including processes ranging from message generation to articulation. By a *comprehensive model of LP* we mean a model that explains how individuals communicate effectively. Such a model would include all processes or components necessary for clear verbal communication. Levelt (1989) provided an example of such a model. In Levelt's (1989) model, speakers must (a) form and maintain a discourse plan, (b) grammatically and phonologically encode the information to be expressed (lexical encoding), (c) monitor both their own and others' speech, (d) edit their speech for errors, and (e) articulate their speech.

The use of a comprehensive model of LP would be helpful for at least three reasons. First, it would allow us to delineate the cognitive processes that operate during LP. For example, a model such as Levelt's (1989) outlines several specific and discrete cognitive processes that play a role in LP. A second advantage to using normal LP models is that they would allow us to combine theories that may each focus on only a subset of LP and thought disorder phenomena. By doing so, we may be able to develop a more complete understanding of thought disorder. Previous thought disorder researchers have implicated disturbances in various components contained in Levelt's LP model,

Deanna M. Barch and Howard Berenbaum, Department of Psychology, University of Illinois at Urbana–Champaign.

This article is based in part on a dissertation submitted by Deanna M. Barch to the Graduate College of the University of Illinois. The research was partially supported by a dissertation research grant awarded by the Graduate College of the University of Illinois.

The contributions of the following dissertation committee members are gratefully acknowledged: Greg Miller, Gary Dell, Barbara O'Keefe, and Larry Hubert. We also wish to thank the staff and patients of Covenant Medical Center in Urbana, Illinois, and Adolph Meyer Zone Center state psychiatric hospital in Decatur, Illinois, for their kind cooperation and support.

Correspondence concerning this article should be addressed to Deanna M. Barch, who is now at Western Psychiatric Institute and Clinic, 3811 O'Hara Street, Pittsburgh, Pennsylvania 15213.

including (a) lexical encoding (e.g., Chapman, Chapman, & Miller, 1964; Spitzer, Braun, Hermle, & Maier, 1993); (b) editing (e.g., Chaika, 1990; Crosson & Hughes, 1987; Maher, 1983; McGrath, 1991); (c) monitoring (e.g., Harvey, 1985; Rochester, 1978); and (d) generating discourse plans (e.g., Hoffman, 1986). Each of these researchers may be correct, but each is focusing on different parts of LP and thought disorder. Using a comprehensive LP model provides a way to integrate multiple theories and to understand how more than one LP deficit could contribute to thought disorder. A third advantage to using a normal LP model is that such a model would permit us to use knowledge gained from basic research on LP to predict, in an empirically and theoretically motivated manner, how disturbances in particular LP components could lead to specific aspects of thought disorder. For example, researchers have proposed that disturbances in the spread of activation in the mental lexicon may be causally related to thought disorder (e.g., Kwapil, Hegley, Chapman, & Chapman, 1990; Spitzer et al., 1993). In many models of normal LP (e.g., Dell, 1986), spreading activation in the mental lexicon enables the retrieval of information during grammatical and phonological encoding. Disturbances in grammatical or phonological encoding may underlie many speech errors that occur in normal speech (e.g., Dell, 1986), as well as the neologisms and jargonistic speech found in some forms of aphasia (Butterworth, 1990; Christman, 1994; Martin, Saffran, Dell, & Schwartz, 1991). Thus, disturbances in grammatical or phonological encoding may be specifically associated with subtypes of thought disorder such as word approximations and neologisms.

Another example of a prediction derived from the literature on normal LP is the hypothesis that incompetent references may be associated with discourse planning deficits. Psycholinguistic theory suggests that the development and maintenance of a discourse plan is integral to adequate reference and coherent language output (Levelt, 1989). Nonpsychiatric individuals with low working memory capacity, who have difficulty maintaining discourse information, also have difficulty producing clear references (Pratt, Boyes, Robins, & Manchester, 1989). Thus, disturbances in discourse planning may be specifically associated with subtypes of thought disorder such as incompetent references.

Four major hypotheses can be distilled from the previous discussion: (a) Thought disorder is a multidimensional construct; (b) multiple causal mechanisms may contribute to diverse dimensions of thought disorder; (c) integrating disparate models of thought disorder is both useful and necessary; and (d) the framework of contemporary models of normal LP provides a way to integrate different thought disorder theories. We believe that significant advances in understanding thought disorder can occur with the investigation of these hypotheses. With this goal in mind, this study was designed to examine which, if any, LP processes are associated with thought disorder. To evaluate whether thought disorder is best conceptualized as a multidimensional phenomenon, we examined whether different LP processes are differentially associated with different facets of thought disorder. Finding that at least one LP process is associated with thought disorder would suggest that the application of normal LP models has the potential to improve our understanding of thought disorder. Finding that LP processes are not equally associated with every facet of thought disorder would suggest that thought disorder is not a unitary phenomenon. Rather, such a finding would be consistent with our hypothesis that thought disorder is a multidimensional phenomenon and that disturbances in different LP processes underlie at least some of the diverse facets of thought disorder.

Method

Participants

Participants were 39 individuals who were hospitalized at the time of their participation in the research project. Using the criteria of the third revised *Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R*; American Psychiatric Association, 1987), 37 participants had diagnoses of schizophrenia and 2 had diagnoses of schizoaffective disorder-manic type. The individuals with schizoaffective disorder were psychotic, but not manic, when they participated. Ninety-two percent of the schizophrenic participants were actively psychotic when they participated. *Psychotic* was defined as displaying one or more of the following symptoms: (a) delusions, (b) hallucinations, or (c) formal thought disorder. Diagnoses were based on the psychotic and mood disorders sections of the Structured Clinical Interview for the *DSM-III-R* (SCID; Spitzer, Williams, Gibbons, & First, 1990) and a review of the participant's clinical records. The interviews were completed by Deanna M. Barch, who had extensive experience using a variety of different structured clinical interviews. Participants were excluded from the study if they scored below 21 on the Mini-Mental State Examination (MMS; Folstein, Folstein, & McHugh, 1975). Only one potential participant was excluded using this criterion. The mean MMS score of the participants was 27.2 (SD = 2.3). Participants were also excluded for (a) substance abuse within the past 6 months; (b) seizure disorders; (c) a history of head trauma; and (d) poor visual acuity (unable to read 12-point font from a distance of $1\frac{1}{2}$ feet [45.7 cm]). Thirty-three percent of the participants were women, and 87% were White. Independent sample t tests indicated no significant gender differences on any of the thought disorder or LP measures. All but five of the participants were receiving some type of psychotropic medication. Of those participants on medication, all were receiving antipsychotics, 38% were receiving lithium, and 59% were receiving antiparkinsonian medication. Of those participants receiving neuroleptic medication, 21 were taking haloperidol, 5 were taking fluphenazine, 2 were taking mesoridazine, and 1 each was taking each of the following; trifluoperazine, clozapine, loxapine, chlorpromazine, molindone, thiothixene, and thioridazine. Additional demographic and clinical characteristics of the participants are shown in Table 1.

To explore the relationship between verbal intelligence, LP, and thought disorder, the interviewer obtained a measure of each participant's verbal intelligence using the Peabody Picture Vocabulary Test—Revised (PPVT-R; Dunn & Dunn, 1981). One participant was unable to complete the PPVT-R. The mean standard score on the PPVT-R for the remaining 38 participants was 84.8 (SD = 20.1, range = 59–133).

Procedure

A research assistant, blind to the participant's diagnosis and current level of symptomatology, conducted the experimental sections of the study. A predetermined written protocol was used to guide the administration of the experimental tasks. Order of task administration was counterbalanced across participants.

Thought Disorder Ratings

To obtain speech samples for thought disorder ratings, participants completed an approximately 10–15-min semistructured interview.

Table 1
Clinical and Demographic Characteristics of Participants

Characteristic	*M*	*SD*	Range
Age	37.1	9.0	20–58
Education (in years)	13.2	3.0	7–21
Age at first hospitalization	22.7	7.7	7–46
Number of previous hospitalizations	8.9	6.9	0–36
Length of current hospitalization (in days)	247.4	443.4	1–1,797
Chlorpromazine equivalents[a]	1,317.5	1,205.9	45–5,100

Note. Daily oral doses of neuroleptics were converted to chlorpromazine equivalents according to guidelines suggested by Davis, Janicak, Linden, Moloney, and Pavkovic (1983). Depot doses were converted to average daily dosages using the guidelines suggested by Baldessarini (1985).
[a] Calculated using only those patients receiving neuroleptics.

This interview consisted of open-ended questions regarding interests and activities. All interviews were completed within 24 hr of the other experimental tasks. The interviews were transcribed by one research assistant and checked for accuracy by two additional research assistants.

The purpose of this study was to determine whether there are differential relationships among thought disorder subtypes and LP deficits, a goal that does not necessitate examining all possible thought disorder subtypes. Furthermore, because of limitations of sample size and power, it was not possible to explore the relationships between LP deficits and all possible thought disorder subtypes. Therefore, we examined only a subset of thought disorder subtypes that we believed would display differential relationships to different LP disturbances. Three trained undergraduate research assistants rated incompetent references according to the criteria outlined by Halliday and Hasan (1976). An *incompetent reference* was defined as a demonstrative, personal, or comparative reference with either an unrecoverable referent or with two or more possible referents. Interrater reliability, measured with an intraclass correlation coefficient, treating raters as random effects and the mean of the raters as the unit of reliability, was .98. To correct for verbosity, the number of incompetent references was divided by the number of words.

Three different trained undergraduate research assistants scored the word-approximation, neologism, tangential response, derailment, and non sequitur response categories from the Scale for the Assessment of Thought, Language, and Communication (TLC; Andreasen, 1986) using the revised definitions described by Berenbaum, Oltmanns, and Gottesman (1985). Scores for each of the categories were derived by summing the number of occurrences per interview and using the corrections suggested by Berenbaum et al. (1985). Interrater reliability, measured using intraclass correlations treating the raters as random effects and the mean of the raters as the unit of reliability, was .75 for word approximations, .75 for neologisms, .69 for tangential responses, .99 for derailments, and .85 for non sequitur responses.

LP Measures

Monitoring task. Monitoring ability was measured using a version of the reality-monitoring task described by Johnson and Raye (1981). Two trials of this task were used to improve reliability and the range of scores. In each trial, participants were shown 16 words, typed individually on notecards, one at a time. The experimenter asked the participant to read 8 of the words aloud ("say" condition) and to imagine reading the other 8 words ("think" condition). The words in each condition were randomly chosen from four lists of eight words that were equal in terms of length and word frequency (based on the Francis & Kucera, 1982, word frequency counts). Participants were then immediately given a list of 24 words, which contained the 16 previously presented words and 8 new words. The participant was asked to indicate whether each word was previously said, previously thought, or a new word. The order of trials and of probe word presentation was counterbalanced across participants. The number of reality-monitoring errors (words incorrectly identified as said that were either new or previously thought) was used as the measure of monitoring ability because, as discussed by Harvey (1985), these types of monitoring errors are most likely to contribute to referential problems. Internal consistency was assessed by computing alpha (.58).

Grammatical–phonological encoding task. Both research and theory suggest that speech errors, particularly nonword errors, at least partially reflect disturbances in grammatical or phonological encoding (e.g., Dell, 1986). Therefore, a version of the speech error induction technique developed by Baars and Motley (Baars & Motley, 1974; Motley & Baars, 1976) was used as the measure of grammatical–phonological encoding. In this task, participants saw a list of word pairs presented on a computer screen. Participants read most of the word pairs silently but were cued to read certain word pairs aloud. The cued word pairs were either filler trials or experimental trials. For experimental trials, the cued word pairs were target pairs designed to elicit either real-word or nonword anticipations, exchanges, or perseverations. An example of a real-word slip is *PEG BET* > > *BEG PET*. An example of a nonword slip is *JAB RUB* > > *RAB JUB*. Each target word pair was preceded by three interference word pairs (e.g., *BOMB PAN, BULL PINES, BELL PUN*) that were more phonologically related to the target error (e.g., *BEG PET*) than to the correct target word pair (e.g., *PEG BET*). The total word list contained 60 target word pairs, 180 filler word pairs, 180 interference word pairs, and 48 nontarget cued word pairs. All of the stimuli, other than 20 newly created word pairs (10 word outcome and 10 nonword outcome) were taken from Dell (1986; 1990). Of the 60 target word pairs, 30 were designed to create real-word slips and 30 were designed to create nonword slips. Stimuli were randomly intermixed and presented in three blocks of 156 trials each, with a break between each block.

Word pairs were shown to participants one at a time in 12-point uppercase letters in the middle of a computer screen. The word pairs were presented for 1,200 ms, and the noncued word pairs were followed by 200 ms of blank screen. The cued word pairs were followed by a row of question marks instead of a blank screen. Participants were told to say the most recent word pair aloud when they saw the row of question marks. Participants were given 1,000 ms to say the cued word pairs, a deadline signaled by a tone. Participants were first given practice trials during which they were urged to speak faster if they missed the 1,000 ms deadline. Participants' verbal responses were tape recorded and four raters independently rated their responses for speech errors. An error was counted if three out of four raters coded it. Nonword errors were used as the measure of grammatical–phonological encoding because nonword errors presumably reflect a greater disturbance in lexical encoding than do word errors (Dell, 1986) and thus may provide a more sensitive and specific measure of grammatical–phonological encoding disturbances. Interrater reliability, measured using an intraclass correlation coefficient treating raters as random effects and the individual rater as the unit of reliability, was .83 for nonword errors. Internal consistency was measured by computing alpha (.77) using the number of nonword errors from each of the three blocks as items.

Discourse planning task. A task was designed to measure the ability to select information relevant to the discourse topic. Participants were given five story topics, one at a time. For each story topic, the participants were shown 20 pictures. The participants were asked to pick out the five pictures that best told the story and to put the pictures in the order that told their story. For each participant, the stories and the pictures were presented in a different random order.

The pictures and story topics were selected after extensive pilot testing with a range of stories and pictures (Barch, 1993). During this pilot testing, undergraduate participants rated the relevance of each of the various pictures to the story topics using a 5-point Likert-type scale, with 1 = *very relevant to the story* and 5 = *completely unrelated to the story*. The interrater reliability of these ratings, using intraclass correlation coefficients treating the raters as random effects and the mean of the raters as the unit of reliability, was .96. As a result of this pilot testing, each picture was assigned a score from 1 to 5 denoting its relevance to telling its associated story topic. The pictures used in the experimental task for each story topic were chosen to represent the full range of possible relevance scores. For each of the five story topics, a measure of the participant's performance for that story was derived by summing the relevance scores for each of their chosen pictures. Thus, each participant was given five relevance scores, one for each story. The overall measure of discourse planning ability for each participant was the sum of these five relevance scores. The internal consistency of this measure was assessed by computing alpha (.69) using the separate relevance scores for each of the five stories as items.

In summary, the three measures of LP processes examined in the data analyses were (a) number of nonword errors on the speech error induction task (grammatical–phonological encoding); (b) the number of reality-monitoring errors (monitoring); and (c) the total relevance score from the discourse planning task (discourse planning). On all of these measures, higher scores indicate worse performance.

Results

We began by examining the correlations among the different thought disorder measures and among the different LP measures. Because the distributions of the thought disorder measures were positively skewed, we used Spearman rank-order correlations instead of parametric correlations for analyses involving thought disorder measures. Most researchers conceptualize word approximations and neologisms as variants of the same phenomenon (e.g., Andreasen, 1986). Consistent with this conceptualization, word approximations and neologisms were highly correlated ($r = .72$, $p < .001$, one-tailed). Therefore, a single score for these two subtypes was created by summing their standardized scores ($\alpha = .89$). Previous research has shown that derailments and non sequiturs are strongly associated (e.g., Berenbaum et al., 1985). We also found that derailments and non sequiturs were correlated ($r = .35$, $p < .05$, one-tailed). Therefore, a single score for these two subtypes was created by summing their standardized scores ($\alpha = .84$). Thus, the four thought disorder scores examined in the remaining data analyses were as follows: (a) word approximations–neologisms; (b) derailments–non sequiturs; (c) incompetent references; and (d) tangential responses. Of these four thought disorder measures, only incompetent references and derailments–non sequiturs were significantly correlated ($r = .30$, $p < .05$, one-tailed). The finding that not all thought disorder subtypes are strongly associated is consistent with the hypothesis that thought disorder is multidimensional.

None of the LP measures were significantly correlated (average $r = .07$, range = $-.07$–$.20$). This result is consistent with the premise that each of these tasks is measuring a different aspect of LP.

Next, we examined the correlations between the thought disorder subtypes and the LP tasks. To reduce the risk of Type 1 errors, we used a Bonferroni correction for the family of correlations involving each LP measure, leading to a corrected significance level of .0125 for each individual correlation. As we had hypothesized, the LP measures were not all equally associated with all thought disorder subtypes. Instead, as can be seen in Table 2, specific LP measures were related to different thought disorder subtypes.

No standard methods are available with which to determine the statistical significance of differences between correlated Spearman rank-order correlations. Thus, to examine the differences between the correlations presented in Table 2, we generated, through Monte Carlo simulations,[1] the probabilities of differences as large or larger than we found occurring under the null hypothesis (e.g., Ho: $r_{xy} = r_{xz}$). To begin, increased referential errors were strongly and specifically associated with worse discourse planning. This specificity was demonstrated in two ways. Incompetent references were significantly more highly correlated with discourse planning than with either grammatical–phonological encoding ($p < .05$) or monitoring ($p < .01$). In addition, discourse planning was significantly more highly correlated with incompetent references than with word approximations–neologisms ($p < .05$), derailments–non sequiturs ($p < .05$), or tangential responses ($p < .05$).

Increased word approximations–neologisms, but none of the other thought disorder subtypes, were strongly and specifically associated with worse grammatical–phonological encoding. Word approximations–neologisms were significantly more highly correlated with grammatical–phonological encoding than with either discourse planning ($p < .001$) or monitoring ($p < .001$). In addition, grammatical–phonological encoding was significantly more highly correlated with word approximations–neologisms than with either derailments–non sequiturs ($p < .05$), or tangential responses ($p < .0001$). There was also a trend for grammatical–phonological encoding to be more highly correlated with word approximations–neologisms than with incompetent references ($p = .06$).

Increased derailments–non sequiturs were associated with worse monitoring performance, although this correlation was not significant after application of the Bonferroni correction.

[1] The null hypothesis (Ho: $r_{xy} = r_{xz}$) implies that the rankings for y and z for a particular participant are identical. If this null hypothesis is correct, one should be able to randomly permute the y and z rankings for each participant and still obtain differences between correlations similar to those we found. Therefore, the statistical significance of our results can be examined by determining the probability under Ho of obtaining differences, in the predicted direction, as large or larger than we found. To generate these probabilities under Ho, we conducted the following simulations for each pair of correlated Spearman rank-order correlations that we wished to compare. For example, to compare the correlation (r_{xy}) between incompetent references (x) and discourse planning (y) with the correlation (r_{xz}) between incompetent references (x) and monitoring (z), we randomly permuted the rankings for discourse planning (y) and monitoring (z) for each participant. We then recalculated two Spearman rank-order correlations (r_{xy}^* and r_{xz}^*) using the permuted rankings and obtained the difference between the two correlations ($r_{xy}^* - r_{xz}^*$). We did this procedure 50,000 times for each pair of correlations and calculated how likely we were to get a difference, under the null hypothesis, as large or larger than we found using the nonpermuted original rankings.

Table 2
Correlations Between Language Production and Communication Disorder Measures

Communication disorder measures	Language production measures: Grammatical–phonological encoding	Monitoring speech	Discourse planning	*M*	*SD*
Neologisms–word approximations	.43**	.04	.01	0.04	2.04
Derailment–non sequitur responses	.01	.32*	.13	0.05	1.95
Incompetent references	.11	.01	.49**	0.01	0.01
Tangential responses	−.02	−.01	.01	0.19	0.35
M	.42	8.59	17.39		
SD	.89	3.58	1.90		

Note. The thought disorder scores were corrected for verbosity, which influenced the low means.
* $p < .05$, one-tailed. ** $p < .01$, one-tailed.

Tangential responses were not significantly associated with any of the LP measures.

Next, we examined the possibility that the links between LP performance and thought disorder subtypes merely reflect a relationship between thought disorder and general verbal intelligence. To explore this possibility, we examined the correlations between the PPVT–R scores and the thought disorder and LP measures. To protect against Type 1 errors, we used a Bonferroni correction to insure a familywise error rate of .05 for the correlations involving the LP measures (per correlation significance level of .0125) and the correlations involving the thought disorder measures (per correlation significance level of .0167). None of these correlations were statistically significant (average $r = -.11$). This finding suggests that the relationships we found between certain thought disorder subtypes and LP performance do not merely reflect associations between thought disorder and general verbal intelligence.

Antipsychotic medications may help reduce thought disorder, potentially decreasing the range of thought disorder scores and thereby increasing the difficulty of detecting an association between thought disorder subtypes and LP processes. The use of correlational analyses with medication dosage levels to examine medication effects on dependent variables is inadequate for several reasons (e.g., Blanchard & Neale, 1992). However, as a preliminary examination of this issue, we examined the relationship between medication dosage, in chlorpromazine equivalents, and the measures of thought disorder and LP. As with all of the previous analyses, we used Bonferroni corrections to control for Type 1 errors. None of these correlations were statistically significant (average $r = -.06$).

Discussion

The goal of this study was to examine the pattern of relationships between LP processes and thought disorder subtypes. Our results indicate that more than one LP process is associated with thought disorder. Specifically, worse performance on all three LP tasks were related to thought disorder. In addition, our results indicate that, as hypothesized, different LP components are differentially associated with different thought disorder subtypes. These differential relationships were demonstrated in two ways. First, not all LP measures were equally associated with all thought disorder subtypes. For example, discourse planning performance was related only to incompetent references and grammatical–phonological encoding was related only to word approximations–neologisms. Second, not all thought disorder subtypes were equally associated with all LP measures. For example, incompetent references were associated only with worse discourse planning performance and word approximations–neologisms were associated only with worse grammatical–phonological encoding performance.

Our results are correlational in nature and thus do not prove that LP deficits play a causal role in thought disorder. However, our results have several implications for future thought disorder research attempting to determine such causal relationships. First, finding that several LP processes are associated with thought disorder suggests that applying normal LP models can improve our understanding of thought disorder. Normal LP models provide a way to integrate thought disorder models and to understand whether and how different LP deficits can lead to different types of language disturbances. In particular, normal LP models may help specify the proximal cognitive mechanisms (e.g., grammatical–phonological encoding) that may contribute to specific thought disorder subtypes. Second, our results indicate that we need to take into account deficits in multiple LP components to fully understand thought disorder. Several LP processes were differentially associated with thought disorder subtypes. This finding attests to the benefits of combining models of thought disorder, each of which focuses on different aspects of LP. Third, our results indicate that it is both necessary and useful to distinguish between different facets of thought disorder. We found that different thought disorder subtypes were differentially associated with different LP processes, providing further evidence that thought disorder is multidimensional.

As we had predicted, our results indicate that worse grammatical–phonological encoding performance was strongly associated with word approximations and neologisms. It should be pointed out that, as a group, the schizophrenic participants produced a percentage of nonword errors (1.4%) similar to that

found in normal participants (e.g., 1.9%; Dell, 1988). However, those schizophrenic participants who displayed word approximations–neologisms produced a much higher percentage of nonword errors (8.3%). Our results are correlational and therefore may not reflect a causal relationship. However, we hypothesize that word approximations and neologisms reflect a disturbance in the ability to selectively access appropriate items during lexical encoding. It has been hypothesized that abnormally fast or far-reaching activation of associations (Spitzer et al., 1993)or a failure to inhibit inappropriate associations (Cohen & Servan-Schrieber, 1992) or both may contribute to language disturbances in schizophrenia. Research on normal LP provides a way to understand how such disturbances could contribute to word approximations and neologisms. In a spreading activation model of normal lexical access, the level of activation determines which item is selected for production. A disturbance in the spread or inhibition of activation could result in inappropriate lexical items being more activated than appropriate lexical items. For example, while attempting to retrieve a target lexical item such as *treehouse*, related lexical items such as *bush* and *cottage* may receive inappropriate amounts of activation. If *bush* and *cottage* were to be more strongly activated than *treehouse,* then *bush* and *cottage* may be chosen for production, leading to a word approximation such as *bushcottage.*

Our results suggest that discourse planning deficits may be associated with incompetent references. It is possible that incompetent references reflect difficulties using, integrating, and maintaining the complex information necessary to guide coherent speech. More specifically, incompetent references may reflect difficulties maintaining and using a discourse plan to select information that is appropriate for the discourse, and hence understandable to the listener.

Derailments–non sequiturs were associated with monitoring performance. However, because this association was not statistically significant after application of a Bonferroni correction, it should be interpreted cautiously. This finding is consistent with previous research indicating a relationship between global measures of thought disorder (whose composition included derailments) and reality-monitoring deficits (e.g., Harvey, Docherty, Serper, & Rasmussen, 1990; Harvey & Serper, 1990). The association between derailments–non sequiturs and monitoring errors may reflect difficulty tracking previous discourse information. Difficulty tracking information about the previous discourse context could lead individuals to produce speech that is unrelated to the previous context. Inspection of the raw data revealed that there was 1 participant who produced a large number of derailments and non sequiturs, but who performed relatively well on the monitoring task. This finding indicates that schizophrenic individuals can produce derailments and non sequiturs in the absence of disturbances in monitoring. Thus, even if poor monitoring can contribute to derailments–non sequiturs, it clearly is not necessary for derailments–non sequiturs to occur. It is also possible that improved measures of the complex skills involved in LP monitoring may clarify its association with derailments–non sequiturs. The reality-monitoring task, which measures memory for single, externally presented words, may not fully tap the complex skills underlying LP monitoring. Additional research is needed to improve our assessment of LP monitoring and to clarify the relationship between monitoring and derailments–non sequiturs, as well as to identify additional factors that contribute to the production of derailments and non sequiturs.

Our results indicated that not all thought disorder subtypes were strongly associated. We believe these findings are consistent with the hypothesis that thought disorder is multidimensional, although the relationship between the correlational level of indicators and dimensional structure is complicated. However, some may be surprised by the relatively low correlations that we found among some of the thought disorder measures. Therefore, we compared our results with those of previous studies. Consistent with previous research, we found a positive correlation between derailments and non sequiturs (Berenbaum et al., 1985) and a positive correlation between derailments–non sequiturs and incompetent references (Harvey & Brault, 1986). In a previous study, we (Berenbaum & Barch, in press) found similar low correlations between word approximations–neologisms and other thought disorder subtypes. We were unable to find other studies that reported correlations between either word approximations or neologisms and other subtypes of thought disorder. Andreasen (1979), Harvey and Brault (1986), and Oltmanns, Murphy, Berenbaum, and Dunlop (1985) all found significant correlations between tangential responses and derailment, whereas we did not. In addition, Harvey and Brault (1986) found a significant correlation between tangential responses and incompetent references, whereas we did not. There may be two reasons for these differences. First, the previous studies (Andreasen, 1979; Harvey & Brault, 1986; Oltmanns et al., 1985) all used the original definitions of derailment and tangential responses developed by Andreasen, whereas we used the definitions used by Berenbaum et al. (1985). The changes made by Berenbaum et al. (1985) to the TLC categories were intended to produce narrower, more specific categories and to prevent the same phenomena (e.g., replies with little or no connection to the question) from being rated in more than one category (e.g., derailment and tangential responses). Thus, it is not surprising that more narrowly defined categories are less strongly correlated than are more broadly defined categories. Second, both Andreasen (1979) and Harvey and Brault (1986) made live ratings of the TLC thought disorder subtypes, whereas we made ratings from transcribed interviews. Live ratings may be much more susceptible to halo effects, leading to inflated correlations among thought disorder subtypes.

We found no significant association between antipsychotic medication dosage and any of the thought disorder or LP variables. As discussed previously, correlational analyses are not adequate for examining medication effects on dependent variables (e.g., Blanchard & Neale, 1992). Thus, further research using both medicated and unmedicated samples is necessary to evaluate the effects of antipsychotic medication on the relationships between thought disorder and LP processes. The most likely effect of antipsychotic medication is to improve thought disorder and potentially LP disturbances. Such improvement would reduce the variability of these measures and could impair our ability to detect relationships between thought disorder and LP disturbances. It is possible that the relationships between thought disorder and LP measures would be even stronger

among unmedicated samples of participants than they were in our medicated sample.

The findings of this study suggest several avenues for future research and theory development. First, further research is needed to determine whether LP deficits play a causal role in thought disorder. This question could be explored by examining whether factors that impair specific LP processes also increase the production of specific thought disorder subtypes. For example, future work should examine whether factors that impair grammatical and phonological encoding, such as increasing speaking rate, also increase the production of word approximations and neologisms. We have used a similar approach in previous work by examining the effects of reducing working memory capacity on the amount and complexity of speech (Barch & Berenbaum, 1994). Second, further research is needed to examine which other LP components (if any) are related to thought disorder. Third, further research is needed to understand the differential relationships among LP components and thought disorder subtypes. For example, disturbances in information processing mechanisms such as selective attention or working memory may also play a role in thought disorder (Cohen & Servan-Schreiber, 1992). It is our view that such cognitive deficits exert their impact on thought disorder by exerting an influence on various aspects of LP. Thus, future theories of thought disorder should integrate models of information processing or brain functioning with models focusing on deficits specific to LP. This integration could be aided by exploring how disturbances in information processing or particular brain systems impair various aspects of LP and thus contribute to specific subtypes of thought disorder. By developing comprehensive models that link biological functioning with cognitive mechanisms and behavioral disturbances, we can acquire a much richer understanding of the pathogenesis of schizophrenia.

References

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

Andreasen, N. C. (1979). Thought, language, and communication disorders: II. Diagnostic significance. *Archives of General Psychiatry, 36,* 1325–1330.

Andreasen, N. C. (1982). Should the term "thought disorder" be revised? *Comprehensive Psychiatry, 23,* 291–299.

Andreasen, N. C. (1986). The scale for assessment of thought, language and communication (TLC). *Schizophrenia Bulletin, 12,* 473–482.

Baars, B. J., & Motley, M. T. (1974). Spoonerisms: Experimental elicitation of human speech errors. *Journal of Supplement Abstract Service, 4* [Catalog of Selected Documents in Psychology].

Baldessarini, R. J., (1985). *Chemotherapy in psychiatry.* Cambridge, MA: Harvard University Press.

Barch, D. (1993). *Communication disorder and language production in schizophrenia.* Unpublished dissertation, University of Illinois at Urbana-Champaign.

Barch, D., & Berenbaum, H. (1994). The relationship between information processing and language production. *Journal of Abnormal Psychology, 103,* 241–250.

Berenbaum, H., & Barch, D. (in press). The categorization of thought disorder. *Journal of Psycholinguistic Research.*

Berenbaum, H., Oltmanns, T. F., & Gottesman, I. I. (1985). Formal thought disorder in schizophrenics and their twins. *Journal of Abnormal Psychology, 94,* 3–16.

Blanchard, J., & Neale, J. M. (1992). Medication effects: Conceptual and methodological issues in schizophrenia research. *Clinical Psychology Review, 12,* 345–361.

Butterworth, B. (1990). Lexical access in speech production. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 108–135). Cambridge, MA: MIT Press.

Chaika, E. (1990). *Understanding psychotic speech: Beyond Freud and Chomsky.* Springfield, IL: Charles C. Thomas.

Chapman, L. J., Chapman, J. P., & Miller, G. A. (1964). A theory of verbal behavior in schizophrenia. In B. Maher (Ed.), *Progress in experimental personality research* (Vol. I, pp. 49–77). New York: Academic Press.

Christman, S. S. (1994). Target-related neologism formation in jargonaphasia. *Brain and Language, 46,* 109–128.

Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex, and dopamine: A connectionist approach to biology and behavior in schizophrenia. *Psychological Review, 99,* 45–77.

Crosson, B., & Hughes, C. W. (1987). Role of the thalamus in language: Is it related to schizophrenic thought disorder? *Schizophrenia Bulletin, 13,* 605–621.

Cutting, J., & Murphy, D. (1988). Schizophrenic thought disorder: A psychological and organic interpretation. *British Journal of Psychiatry, 152,* 310–319.

Davis, J., Janicak, P., Linden, R., Moloney, J., & Pavkovic, I. (1983). Neuroleptics and psychotic disorders. In J. T. Coyle & S. J. Enna (Eds.), *Neuroleptics: Neurochemical, behavioral, and clinical perspectives* (pp. 15–64). New York: Raven Press.

Dell, G. S. (1986). Spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

Dell, G. S. (1988). The retrieval of phonological forms in production: Test of predictions from a connectionist model. *Journal of Memory and Language, 27,* 124–142.

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes, 5,* 313–349.

Dunn, L. M., & Dunn, L. M. (1981). *The Peabody Picture Vocabulary Test—Revised.* Circle Pines, MN: American Guidance Service.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189–198.

Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar.* Boston: Houghton Mifflin Company.

Gray, J. A., Feldon, J., Rawlins, J. N. P., Hemsley, D. R., & Smith, A. D. (1991). The neuropsychology of schizophrenia. *Behavioral and Brain Sciences, 14,* 1–84.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in spoken and written English.* London: Longmans.

Harrow, M., Lanin-Kettering, I., & Miller, J. G. (1989). Impaired perspective and thought pathology in schizophrenic and psychotic disorders. *Schizophrenia Bulletin, 15,* 605–623.

Harvey, P. D. (1985). Reality monitoring in mania and schizophrenia: The association of thought disorder and performance. *The Journal of Nervous and Mental Disease, 173,* 67–73.

Harvey, P. D., & Brault, J. (1986). Speech performance in mania and schizophrenia: The association of positive and negative thought disorder and reference failures. *Journal of Communication Disorders, 19,* 161–173.

Harvey, P. D., Docherty, N. M., Serper, M. R., & Rasmussen, M. (1990). Cognitive deficits and thought disorder: II. An 8-month follow-up study. *Schizophrenia Bulletin, 16,* 147–156.

Harvey, P. D., & Serper, M. R. (1990). Linguistic and cognitive failures in schizophrenics: A multivariate analysis. *The Journal of Nervous and Mental Disease, 178,* 487–494.

Hoffman, R. E. (1986). Verbal hallucinations and language processes in schizophrenia. *The Behavioral and Brain Sciences, 9,* 503–548.

Johnson, M. D., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88,* 67–85.

Kwapil, T. R., Hegley, D. C., Chapman, L. J., & Chapman, J. P. (1990). Facilitation of word recognition by semantic priming in schizophrenia. *Journal of Abnormal Psychology, 99,* 215–221.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Maher, B. A. (1983). A tentative theory of schizophrenic utterances. *Progress in Experimental Personality Research, 12,* 1–52.

Maher, B. A. (1991). Language and schizophrenia. In S. R. Steinhauer, J. H. Gruzelier, & J. Zubin (Eds.), *Handbook of schizophrenia: Vol. 5. Neuropsychology, psychophysiology, and information processing* (pp. 437–464). Amsterdam: Elsevier.

Martin, H., Saffran, E. M., Dell, G. S., & Schwartz, M. F. (1991, October). *On the origin of paraphasic errors in deep dysphasia: Simulating error patterns in deep dysphasia.* Paper presented at the Deep Dyslexia II meeting, London.

McGrath, J. (1991). Ordering thoughts on thought disorder. *British Journal of Psychiatry, 158,* 307–316.

Motley, M. T., & Baars, B. J. (1976). Laboratory induction of verbal slips: A new method for psycholinguistic research. *Communication Quarterly, 24,* 28–34.

Oltmanns, T. F., Murphy, R., Berenbaum, H., & Dunlop, S. R. (1985). Rating verbal communication impairment in schizophrenia and affective disorders. *Schizophrenia Bulletin, 11,* 292–299.

Pratt, M. W., Boyes, C., Robins, S., & Manchester, J. (1989). Telling tales: Aging, working memory, and the narrative cohesion of story retellings. *Developmental Psychology, 4,* 628–635.

Rochester, S. R. (1978). Are language disorders in acute schizophrenia actually information processing problems? *Journal of Psychiatric Research, 14,* 275–283.

Spitzer, M., Braun, U., Hermle, L., & Maier, S. (1993). Associative semantic network dysfunction in thought-disordered schizophrenic patients: Direct evidence from indirect semantic priming. *Biological Psychiatry, 34,* 864–877.

Spitzer, R. L., Williams, J. B. W., Gibbon, M., & First, M. B. (1990). *Structured Clinical Interview for DSM-III-R—Patient Edition (SCID-P, Version 1.0).* Washington, DC: American Psychiatric Press.

Received June 27, 1994
Revision received May 23, 1995
Accepted July 19, 1995 ■