Neuropsychology
2005, Vol. 19, No. 3, 371–380

DOI: 10.1037/0894-4105.19.3.371

Category Content and Structure in Schizophrenia: An Evaluation Using the Instantiation Principle

Brita Elvevåg
National Institute of Mental Health

Gert Storms
University of Leuven

Evan Heit
University of Warwick

Terry Goldberg
National Institute of Mental Health

In numerous studies, researchers have suggested anomalies in semantics in patients with schizophrenia. In this study, the authors addressed whether one such anomaly might reflect a difference in knowledge content or in the structure or organization of this information. Using a category member production task and a typicality rating task, the authors assessed knowledge content and found that patients' and control participants' knowledge about categories of foods and animals was very similar. In terms of structure, their findings from a mathematical model of category judgment (the instantiation model; E. Heit & L. W. Barsalou, 1996) revealed a similar category structure in patients and control participants. In conclusion, the authors suggest that the content and organization of categories in patients with schizophrenia is similar to that in healthy control participants.

Keywords: exemplars, prototypes, categories, semantic memory, schizophrenia

The study of semantics deals with words as signs (symbols) and language that refer to entities in the world. The postulated storage and organization of these meanings is often referred to as *semantic memory*. There is a growing research interest in schizophrenia concerning which aspects of semantic memory are preserved and which are dysfunctional and how these might relate to the cognitive deficit profile of individuals with schizophrenia, as well as to their symptoms. Specifically, we sought to examine the content and structure of this information (within the same experimental paradigm) in natural language categories and to evaluate how normal this is in patients with schizophrenia.

In a variety of studies, researchers have explored knowledge content and structure in schizophrenia but have used different paradigms to examine these components of semantic memory. For example, in an examination of semantic memory content in patients with schizophrenia, Tamlyn et al. (1992) used a simple semantic sentence verification task to index real-world knowledge. The task was to state, as quickly as possible, whether each of 50 spoken sentences (e.g., "Rats have teeth" and "Desks wear clothes") are true or false. They found that patients' verification times were significantly slower than that of control participants from normative data and that patients made a surprising (given how easy the task is) number of errors. In another study focusing predominantly on structure of knowledge, researchers used a comprehensive semantic test battery (designed by Hodges, Salmon, & Butters, 1992) that systematically varies multiple input and output modalities to build a profile of the status of semantic knowledge that is independent of any one modality used to access or express it. The researchers probed knowledge of living and humanmade items through tasks of category fluency, picture naming, sorting of picture cards into different categories, and word-to-picture matching, and finally, they required participants to generate defining features of items. MacKay et al. (1996) reported that patients with schizophrenia performed worse on almost all of the semantic subtests than did healthy control participants, although in the word-to-picture matching subtest, the group differences were less pronounced. Of interest is that patients' performance was, on average, intermediate between that of healthy individuals and patients with mild to moderate Alzheimer's disease.

Other data that may be broadly viewed as examining knowledge structure in schizophrenia are category fluency data, in which a participant generates as many exemplars of a certain category (e.g., an animal) as possible in a specified time. Although poor performance on this category fluency task by patients with schizophrenia has been used to argue that semantic memory is organized differently (i.e., suboptimally for access) in these individuals (e.g., Sumiyoshi et al., 2001; but see Elvevåg & Storms, 2003, and Storms, Dirikx, Saerens, Verstraeten, & De Deyn, 2003, for an evaluation of the statistical problems with this data-fitting approach to patient data), other data suggest that the clustered output structure on a verbal category fluency task is similar to that of control participants (Elvevåg, Fisher, Gurd, & Goldberg, 2002). Moreover, an indirect assessment of the amount of information on specific categories in the semantic store was found to be equivalent in patients and control participants (i.e., the lexicon size is intact; Elvevåg, Weinstock, Akil, Kleinman, & Goldberg, 2001; see also Allen, Liddle, & Frith, 1993). These data have been interpreted within the framework provided by the "storage versus access" debate in semantics as reflecting a problem in the retrieval of words from the lexicon (i.e., the semantic store; e.g., Allen et al., 1993).

Brita Elvevåg and Terry Goldberg, Clinical Brain Disorders Branch, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland; Gert Storms, Psychology Department, University of Leuven, Leuven, Belgium; Evan Heit, Department of Psychology, University of Warwick, Warwick, United Kingdom.

Correspondence concerning this article should be addressed to Brita Elvevåg, Clinical Brain Disorders Branch, National Institute of Mental Health, National Institutes of Health, Building 10, Room 4S235, MSC 1379, Bethesda, MD 20892. E-mail: elvevaab@intra.nimh.nih.gov

In a similar vein, the idea of a problem in accessing knowledge versus a problem in knowledge structure has been examined in semantic priming studies, which are based on the idea that semantic memory is organized as a network (Collins & Quillian, 1969). This process has been indexed with the use of tasks in which target words are preceded by a prime (a cue), which, if related to the target, speeds the recognition of the target. These tasks have been used to demonstrate deviations in the associational network, or aberrant associations between words, in patients with schizophrenia (e.g., Goldberg et al., 1998; Goldberg, Dodge, Aloia, Egan, & Weinberger, 2000; Manschreck et al., 1988; Ober, Vinogradov, & Shenaut, 1995; Spitzer, 1997; Spitzer, Braun, Hermle, & Maier, 1993; Spitzer et al., 1994; Vinogradov, Ober, & Shenaut, 1992). They have also been variously interpreted as suggesting that there is a problem of concepts appropriately activating each other.

Our own research takes these previous studies as a starting point. Given the previous indications that there are some semantic deficits in patients with schizophrenia, we sought to pinpoint this deficit. In particular, we explored to what extent such deficits are related to the content of information that is assessed or to the structure and organization of this knowledge. Our main goal was to compare patients with schizophrenia to healthy control participants in terms of their semantic knowledge about natural language categories (foods and animals) and to assess whether any group differences were driven by differences in content or structure.

First, to examine content of knowledge, we used two tasks: a category member production task and a typicality rating task. These tasks allowed us to assess whether patients and control participants are alike in terms of naming category members (e.g., naming *dog* as an example of *mammal*) and judging their typicality to a superordinate (e.g., rating how typical a dog is to the superordinate *animal*). We predicted that if patients have different knowledge about natural language categories in terms of content, then exemplars generated in response to a specific category cue (e.g., *mammal*) would have a substantially different distribution from that of control participants. Likewise, a difference in content of knowledge should lead to differences in typicality ratings for category members (e.g., *dog*). Second, to examine structure of knowledge, we applied the instantiation model (Heit & Barsalou, 1996) to data from the two groups. The instantiation model predicts a systematic relationship between two levels of categorical structure in terms of statistical regularities between typicality ratings at the two taxonomic levels (e.g., the instantiation *dog* and the superordinate *animal*). The instantiation model has been validated previously for healthy populations for food and animal categories, and with its use, researchers are able to predict the statistical distributions (in terms of means and standard deviations) of typicality ratings for the categories (e.g., *mammal*) from the distributions of typicality ratings on the instantiations (e.g., *dog*; see Heit & Barsalou, 1996, for details, as well as the Results and Discussion section below). Heit and Barsalou (1996) concluded that the good fit of the instantiation model to these data suggests that people were consistent between the two sets of judgments (see also De Wilde, Vanoverberghe, Storms, & De Boeck, 2003). Applying the instantiation model to data from patients with schizophrenia should be informative about the consistency of patients' category representation. A poor fit of the instantiation model for patients with schizophrenia would suggest that their knowledge is not as systematically structured or is structured in different ways.

Therefore, in the present study we compared the fit of the instantiation model for patients to the fit for a control group, our goal being to compare patients' pattern of behavior with control participants' pattern of behavior.[1] By no means was it our goal to propose the instantiation model as a diagnostic test or model of schizophrenia. Indeed, it could well be the case that the instantiation model results in a good fit with data from patients with schizophrenia. Such an outcome would not be a null result, because such a good fit of the instantiation model would require a statistically significant relation between dozens of judgments made at two taxonomic levels. Therefore, such an impressive outcome would present a very convincing case for some aspects of patients' semantic category knowledge in terms of content and structure being well preserved (i.e., normal).

In summary, we hypothesized that if patients have different knowledge about natural categories in terms of content, then the exemplars generated in response to a specific category cue would have a substantially different distribution from that of control participants and should lead to differences in typicality ratings for category members. Furthermore, if control participants and patients have similar knowledge about certain categories in terms of content, but patients display less consistently structured knowledge or knowledge that is structured in a different way, the instantiation model would fit the data of patients worse than that of control participants. Indeed, if the content of patients' knowledge is different but otherwise structured systematically in a similar fashion to that of control participants, then the instantiation model could still give a good fit to the patients' data. Furthermore, if the predictions based on the data of the control participants predicted the performance of the patients well and vice versa, then that would provide evidence for very similar semantic structure across the groups.

Method

Baseline Tests

As a baseline assessment of intellectual function, all participants were administered the Wide Range Achievement Test—Revised (WRAT–R; Jastak & Wilkinson, 1984). The WRAT–R is widely used as a putative measure of premorbid intelligence (Goldberg et al., 1995; Kremen et al., 1996; Wiens, Bryan, & Crossen, 1993). To estimate current intelligence, we administered a short form of the Wechsler Adult Intelligence Scale—Revised (WAIS–R; Wechsler, 1981; see also Kaufman, 1990; Missar, Gold, & Goldberg, 1994). The observed substantial drop in intelligence from the estimated premorbid function is often reported in schizophrenia (e.g., see Weickert et al., 2000). Participants were also administered a standard test of current nonverbal or fluid intelligence (g_f): Cattell's Culture Fair Intelligence Test (Cattell, 1971; Institute for Personality and Ability Testing, 1973), Scale 2, Form A. The test involves timed spatial problem-solving tasks.

Participants

There were 21 patients with schizophrenia and 22 control participants in the production task. The data from these participants were collated separately for each group and used for the subsequent rating task, which was administered to a separate sample of 20 patients and 21 control participants

[1] Naturally, the model has limits. For example, the model does not specify how such knowledge will be used in speech.

in the rating task (for characteristics of patient and control samples, see Table 1).[2] These same individuals (patients and control participants) participated in Study 1: Food and Study 2: Animals, which were administered on the same day. Participants with schizophrenia were patients from the National Institute of Mental Health's (NIMH) Neuropsychiatric Research Center at St. Elizabeth's Hospital, Washington, D.C., and inpatients and outpatients from the NIMH's research ward in Bethesda, Maryland. All patients fulfilled criteria of the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM–IV*; American Psychiatric Association, 1994) for schizophrenia or schizoaffective disorder as determined by the Structured Clinical Interview for *DSM–IV* (SCID), with three psychiatrists reaching a consensus diagnosis. Patients generally had multiple hospital admissions due to incomplete responses to conventional treatments. Healthy control volunteers were recruited through the National Institutes of Health's volunteer panel. No participant, control or patient, with a history of traumatic brain injury, epilepsy, developmental disorder, diagnosable current substance or alcohol abuse or dependence, or other known neurologic condition was included in this study. All participants had normal or normal corrected vision. All control participants and outpatients were paid for their participation, and inpatients completed the study as part of their protocol for entering the hospital. The study was approved by NIMH's internal review board, and informed consent was obtained from all participants prior to testing.

Materials and Procedure

For the production groups, the category cues were the nine subordinates of food, namely, *beverage, dairy product, dessert, fish, fruit, meat, poultry, seasoning,* and *vegetable,* and the seven subordinates of animal, namely, *amphibian, bird, fish, insect, mammal, microorganism,* and *reptile*. These same seven subordinates of animal were modified by the adjectives *small* and also *dangerous,* providing a total of 21 animal cues. Participants were individually asked to name the first three instances that came to mind for each category cue. The cues were read to each participant one at a time in a pseudorandom order, and the participants' responses were recorded. Of importance, we note that participants were not prevented from making mistakes, such as producing *crab* as an instantiation of *fish*. Production control participants produced a total of 178 unique instantiations to the 9 food cues and 265 unique instantiations to the 21 animal cues, and production patients produced a total of 185 unique instantiations to the food cues and 269 unique instantiations to the animal cues.

For the control rating group, the stimuli to be rated were the unique instantiations produced by the control production group,[3] and for the patient rating group, the stimuli to be rated were the unique instantiations produced by the patient production group. Additionally, the subordinate categories (i.e., the 9 subordinates of food and 21 subordinates of animal) were also rated by both rating groups. Participants were to rate the food words on how typical a member they were of the food category and how typical the animal words were of the animal category. They were given a clear illustration of the 1–9 scale that they were to use, with higher numbers indicating greater typicality. For example, participants were asked how typical *milk* is for the category *food* and how typical *beverage* is for the category *food*. Note that participants never rated the instantiations on their typicality in the subordinate categories. That is, participants never evaluated the typicality of *milk* in *beverage*. Participants performed the rating task at their own pace.

Results and Discussion

First, we describe the results of the two tasks assessing the content of knowledge, namely, the production task and the typicality rating task. Then we present the application of the instantiation model, which we used to assess the structure of knowledge for the patient group and the control group. Finally, we present the

Table 1
Characteristics of Patient and Control Samples in the Production Task and the Rating Task

Variable	Patient[a] *M*	Patient[a] *SD*	Control[b] *M*	Control[b] *SD*
Production task				
Age (years)*	37.00	10.44	30.14	6.91
WAIS–R IQ*	90.67	14.28	112.71	14.80
WRAT–R IQ	99.67	16.16	107.76	12.25
Cattell IQ*	82.78	19.15	113.90	18.73
Age at 1st hospitalization (years)	21.19	4.19		
Years since 1st hospitalization	15.81	10.25		
Neuroleptic medication	21		0	
Clozapine/olanzapine	9			
Risperidone	5			
High-potency drugs[c]	7			
Anticholinergics	7			
Adjunctives[d]	9			
Rating task				
Age (years)	37.35	8.27	36.00	7.73
WAIS–R IQ*	92.90	9.96	103.57	10.19
WRAT–R IQ	106.25	10.13	102.62	8.86
Cattell IQ*	94.47	18.17	106.80	15.67
Age at 1st hospitalization (years)	20.05	4.66		
Years since 1st hospitalization	17.30	8.49		
Neuroleptic medication	20		0	
Clozapine/olanzapine/quetiapine	13			
Risperidone	6			
High-potency drugs[c]	4			
Anticholinergics	2			
Adjunctives[d]	13			

Note. Independent samples *t* tests did not show any significant differences between patients in the production group and patients in the rating group in age, WAIS–R IQ, WRAT–R IQ, Cattell IQ, age at first hospitalization, and number of years since first hospitalization (in all cases, $p > .05$). For the control sample, there were significant differences between control participants in the production group and control participants in the rating group in age and WAIS–R IQ (in both cases, $p < .05$), although there were not any differences in WRAT–R IQ and Cattell IQ values (in both cases, $p > .1$). WAIS–R = Wechsler Adult Intelligence Scale—Revised; WRAT–R = Wide Range Achievement Test—Revised; Cattell = Cattell's Culture Fair Intelligence Test.
[a] For the production task, $n = 21$ (17 men, 4 women); for the rating task, $n = 20$ (18 men, 2 women). [b] For the production task, $n = 22$ (11 men, 11 women); for the rating task, $n = 21$ (11 men, 10 women). [c] Haloperidol, fluphenazine, and loxapine. [d] Lithium, Depakote, sertraline, lorazepam, venlafaxine, clonazepam, and buspirone.
* $p < .01$ (independent samples *t* test).

[2] Two separate samples of participants took part in the production and rating tasks because of the time lag involved in collecting data from 21 patients on a ward that on average had only 10 patients at any point in time.

[3] Because of an error in the design of the rating sheets for control participants, the word *ant* was omitted. Therefore, we obtained typicality ratings of the word *ant* by testing a further 21 healthy control participants. We gave them three lists of 10 words for which they were to rate the typicality of the items for the categories *animal, small animal,* and *dangerous animal.* The 10 words were selected from the entire production list and spanned the continuum of typicality frequency (as rated by control participants).

results of simulations used to evaluate the sensitivity of the instantiation model to different types of noise in the semantic system.

Study 1: Food

Production task. We first examined the category members produced for the nine subcategories of food. The question of whether the category members produced by patients and control participants differ can be answered by looking at several different indices.

First, we looked at the number of distinct exemplars produced for each of the nine subcategories of food, taking into account only the first instantiation of each participant. A reason for concentrating on the first instantiations only is that these can be considered the most prototypical, because production frequency and typicality correlate significantly (Storms, De Boeck, & Ruts, 2000). For *beverage, dairy product, dessert, fish, fruit, meat, poultry, seasoning,* and *vegetable,* the control participants generated 13, 5, 9, 13, 7, 11, 5, 13, and 12 distinct exemplars, respectively. The corresponding numbers of the patient group were very similar: 15, 7, 11, 10, 6, 11, 6, 12, and 12, respectively. A paired t test did not show a significant difference, $t(8) = -.39$, $p = .71$. Across the nine subcategories, there was a significant correlation between the exemplars produced by the two groups ($r = .86$, $p < .01$). Thus, in the number of distinct exemplars produced for each category, the results were similar for patients and control participants.

Second, we examined to what extent the actual words generated as first instantiations were similar across groups. For the nine subordinate categories, the percentage of instantiations in the patient group that were also generated in the control group varied between 52% and 90%. These percentages are fairly high, given that only the first generated instantiations in a relatively small group ($n = 21$) were compared. Moreover, for seven of nine subordinates, the control participants' most frequent instantiation was also the patients' most frequent instantiation. Occasionally, odd exemplars were generated (such as *eggs* for *poultry*), but these peculiarities were rare in both the patient group and the control group. Overall, there was no evidence that the patients produced significantly less common or more unusual category members (or nonmembers) or responded with a substantially different probability distribution than did the control participants.

Typicality rating task. A first important question relates to how similar the typicality ratings were within each group. It is conceivable that the control participants were highly interindividually consistent (i.e., that they all rated the items in a very similar way) but that the patients differed considerably among each other in the way they rated typicality. Therefore, interindividual consistency within each of the two groups was evaluated with a standard psychometric reliability estimation procedure. To obtain this estimate of the reliability of these mean ratings, we split the two groups at random into two half groups. Next, the mean ratings were calculated within each half group. Then, the interscorer reliability (i.e., interindividual consistency) of the mean ratings was estimated with the split-half correlation, corrected with the Spearman–Brown formula (Lord & Novick, 1968). The correlation between the means in the two half groups of control participants was .90, which resulted in an estimated reliability of .94. The corresponding correlation for the two half groups of patients was .86, which resulted in an estimated reliability of .93. In other words, there was a high degree of agreement within the patient group as well as within the control group, as reflected in the means. The difference between the patients' correlation and the control participants' correlation was not significant ($z = 1.11$). These correlations, as well as the remaining correlations in this section on typicality ratings, were calculated over all instantiations that were rated in both the patient group and the control group. Thus, this finding argues against very idiosyncratic differences among the patients (and also among the control participants) in their typicality ratings.

Next, we conducted analyses comparing the consistency within each group, control and patient, to the consistency between groups. The key question was whether the differences between patients and control participants were greater than the differences within each group considered alone. To examine how closely the ratings of patients resembled those of control participants, we calculated correlations between the half groups of patients and the half groups of control participants over all exemplars that were rated by both groups of patients and control participants. In this way, four different correlations could be calculated. The resulting values were .81, .87, .78, and .82, respectively. None of these values differed significantly from the correlations between the two half groups of patients or from the correlations between the two half groups of control participants. Hence, the between-groups correlations were not lower than the within-group correlations.

We further investigated the homogeneity of the two groups by calculating, within every half group, the standard deviation of the typicality ratings for every instantiation. The reason we investigated the within-item standard deviations of the ratings in detail is because the fit of Heit and Barsalou's (1996) instantiation model, which we discuss below, relies heavily on the within-item variance of the ratings. The correlation between these standard deviations (i.e., one for every instantiation) in the two half groups of control participants was .79 (which results in an estimated reliability of the standard deviations equal to .88). The corresponding correlation for the two half groups of patients was only .61 (which results in an estimated reliability of .76). The difference between the patient correlation and the control correlation was significant ($z = 2.74$, $p < .01$). In other words, the degree of agreement within the control group, as reflected in the standard deviations, was higher than the agreement within the patient group.

In summary, we used the measures of reliability of typicality ratings to address two issues. One issue is whether the two groups were equally consistent. The reliability analyses suggest that control participants and patients were equally consistent in mean typicality ratings, but there was more disagreement in the patient group when the standard deviations of typicality ratings were considered. The second issue is whether a mixed group, half control and half patient, would be less consistent than a pure group, either all control or all patient. If the mixed group is less consistent, it would suggest that there are indeed differences between control participants and patients. However, for mean typicality ratings, we found that the mixed groups were about as consistent as the pure groups. For standard deviations of typicality ratings, we found that the mixed groups were not significantly less consistent than the pure patient groups, which themselves were somewhat inconsistent. Overall, the only evidence for a difference between patients and control participants in typicality ratings was minor: The standard deviation measure was less reliable for patients.

The instantiation model. When participants are asked how typical *beverage* is of the category *food,* the instantiation model assumes that first an example of *beverage* (e.g., *Coca-Cola*) is generated, and then the evaluation of how typical *Coca-Cola* is of the category *food* occurs. The manner in which the model implements this is to correlate the observed ratings of typicalities of *beverage* with the predicted typicalities of *beverage.* These predicted typicalities are derived as follows: If 5 out of 20 participants in the original production task generated *Coca-Cola* in response to the cue *beverage*, then *Coca-Cola* was used 25% of the time to predict how typical *beverage* is of the category *food*. Next, a typicality rating for the instantiation was chosen randomly from the 21 typicality judgments that the participants in the rating task made for it. This was continued until all the typicality rating data entered into the simulation. Overall, a total of 420 participants (21 generation participants × 20 rating participants) were simulated for each subordinate in 20 cycles of 21 simulated participants each, thus ensuring that the relative frequency of different instantiations is maintained in the simulated data.

The key results address the fit of the instantiation model for patients and control participants. Because of consistency within the groups, with the instantiation model we were able to make good predictions of patients' mean typicality ratings for the nine subordinates on the basis of information from the instantiations the patients produced (r = .82; see Figure 1). Likewise, use of the model enabled us to make good predictions of control participants' mean typicality ratings for the nine subordinates on the basis of information from their instantiations (r = .82).[4] Also, looking just at the observed means for the nine subordinates, we see a high correlation (r = .87) between the means for patients and the means for control participants, which again suggests that there was much overlap in content. Furthermore, because there was a large amount of consistency between patients' and control participants' data, it was possible to predict patients' subordinate judgments from control participants' instantiations (r = .79) and control participants' subordinate judgments from patients' instantiations (r = .72).

We next turn to information about standard deviations of the distributions of typicality ratings for the nine subordinate categories. In general, the correlations were lower for standard deviations than for means. Heit and Barsalou (1996) had a similar finding, with an overall fit of .64 for the instantiation model's prediction of standard deviations for foods, and they attributed this lower correlation to a more restricted range for standard deviations and also to the fact that means are inherently more efficient (reliable) estimators than standard deviations. In our study, the key results were a fit of .71 for patients and a fit of .81 for control participants, both of which were higher than Heit and Barsalou's own results. The fits for patients versus control participants did not differ significantly. The other positive correlations reflect that there was some consistency between patients and control participants, although not a perfect degree of consistency. For example, the model was better able to make predictions for patients' subordinates from patients' instantiations (r = .71) than from control participants' instantiations (r = .33).

In summary, patients' and control participants' knowledge about food categories seemed fairly similar. In terms of content, we did not observe significant differences in the category members produced or in average typicality ratings on these category members, although patients displayed somewhat more inconsistency in the standard deviation measures. In terms of structure, the instantiation model, a mathematical model of category structure, was able to predict systematic relations between two levels of categorical structure in means and standard deviations. The instantiation model was successful in predicting both patient and control performance, and the two groups did not differ significantly in terms of how well the instantiation model fit their data. If patients' knowledge was less systematic than control participants' knowl-

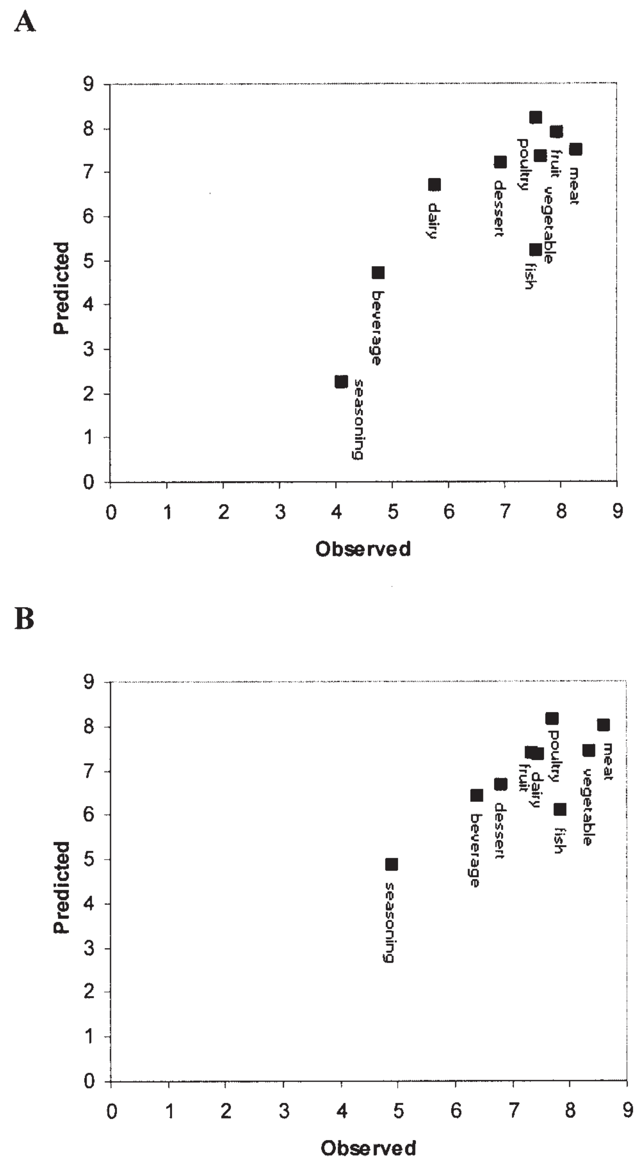

Figure 1. Study 1: Observed and predicted means of typicality ratings for food in control participants (A) and in patients with schizophrenia (B).

[4] For comparison, note that Heit and Barsalou (1996) found a correlation of .89 in a similar study with foods.

edge, or if patients' knowledge was structured in a different way, then the instantiation model should not have fit the data as well.

In Study 2, we attempted to replicate this result using another semantic domain, animals rather than foods. Knowledge of animals and living things is perhaps a more usual domain of study for schizophrenia. In addition, the stimulus set for animals was much more extensive than the stimulus set for foods. For example, by including combined concepts such as *small bird* and *dangerous reptile*, we were able to assess the fit of the instantiation model over 63 subordinates in Study 2, in comparison with 9 subordinates in Study 1.

Study 2: Animals

Production task. Following the procedure in Study 1, we examined the total number of distinct category members for the seven unmodified subcategories of animals (i.e., *amphibian, bird, fish, insect, mammal, microorganism,* and *reptile*) and for the same seven subcategories modified with the adjective *dangerous* (e.g., *dangerous amphibian, dangerous bird*) and the adjective *small* (e.g., *small amphibian, small bird*). This comparison over the 21 subcategories resulted in a mean of 11.5 and 9.9 distinct instantiations for patients and control participants, respectively, on the basis of the first instantiations only. This difference was not significant, $t(20) = 1.26$. For these 21 subordinate categories, the percentage of instantiations in the patient group that were also generated in the control group averaged 60%, which is high given that only the first generated instantiations in a relatively small group ($n = 21$) were compared. Also, for 10 of the 21 subordinates, the most frequent instantiation produced by the control participants was also the most frequent instantiation produced by the patients. These results suggest that over a range of subcategories of animals, including combined concepts, patients and control participants were similar in their performance on the category member production task. These findings replicate that of Study 1 and provide evidence for its generality.

Typicality rating task. Mean ratings over participants were calculated for every instantiation, separately within each group, for all instantiations that were rated in both the patient and the control group. As in Study 1, the interscorer reliability of the mean ratings was estimated. The correlation between the means in the two half groups of control participants was .91, resulting in an estimated reliability of .96, and the corresponding correlation for patients was .79, resulting in an estimated reliability of .89. The difference between the patients' and the control participants' correlation was significant ($z = 7.78$, $p < .01$), indicating a high degree of agreement within the control group and a somewhat lower agreement within the patient group. Correlations were calculated between half groups of patients and half groups of control participants over all exemplars that were rated by both groups, and the resulting values were .88, .88, .87, and .87, respectively. None of these values differed significantly from the correlations between the two half groups of patients nor from the correlations between the two half groups of control participants.

The homogeneity of the two groups was examined by calculating, within every half group, the standard deviation of the typicality ratings for every instantiation. The correlation between these standard deviations (i.e., one for every instantiation) in the two half groups of control participants was .64 (resulting in an estimated reliability equal to .78) and in the two half groups of patients, only .35 (resulting in an estimated reliability of .52). The difference between the patients' and the control participants' correlation was significant ($z = 6.43$, $p < .01$), indicating that the degree of agreement within the control group was larger than that within the patient group.

In summary, there was some difference between control participants and patients in that as a group, patients were somewhat less consistent. However, when mixed groups, half control and half patient, were examined, these mixed groups were not significantly less consistent than the patients themselves. Hence, the difference between control participants and patients was not greater than any differences within the group of patients.

The instantiation model. We examined the correlations between observed and predicted means and standard deviations, respectively, for the two groups. These correlations were taken over 63 items, that is, for the 21 subordinates using typicality ratings in *animal, dangerous animal,* and *small animal* together. In terms of means, the instantiation model was able to predict typicality ratings very well for both patients ($r = .88$) and control participants ($r = .90$; see Figure 2), as it was for standard deviations ($r = .70$ for patients and $r = .74$ for control participants). In all cases, each correlation was significantly different from chance, but the patient and control groups did not differ significantly from each other. In summary, the instantiation model appeared to give a very strong account of the structural relations between the subordinate categories and their instantiations, for both the patient and the control groups.

Simulations

In light of our findings of a similar behavioral pattern in patients and control participants, the question emerges as to how sensitive our approach is in detecting a possible impairment in the semantic domain. It seems very reasonable to assume that, if the patients show semantic deficits, the impairments of individual patients are not exactly identical. Thus, the resulting increased inconsistency in the data from the patient group could resemble increased randomness in the data.

To address the sensitivity of the instantiation model to randomness, we present simulations that demonstrate that adding error to the data results in decreasing fit measures. For illustration purposes, we simulated data for the 63 animal categories because large amounts of data are more informative and representative (than just a few categories), especially with regard to correlations. Naturally, there are many ways that error can be introduced into a semantic system. We describe two different procedures in detail below.

First, we introduced randomness into the instantiation process, the idea being that a person with a semantically disordered system might generate many incorrect instantiations for a given subordinate (e.g., if asked to generate a *mammal,* the person names a dangerous fish, such as *piranha*). Two hundred data sets were run for four different percentages of random data (20%, 25%, 50%, and 75%). For instance, with 25% random data, a fourth of the instantiations were randomly shuffled. Thus, for these shuffled instantiations, the corresponding typicalities on which the calculations of the instantiation model were based were derived from incorrect animals (e.g., typicalities from *cat* could be used to

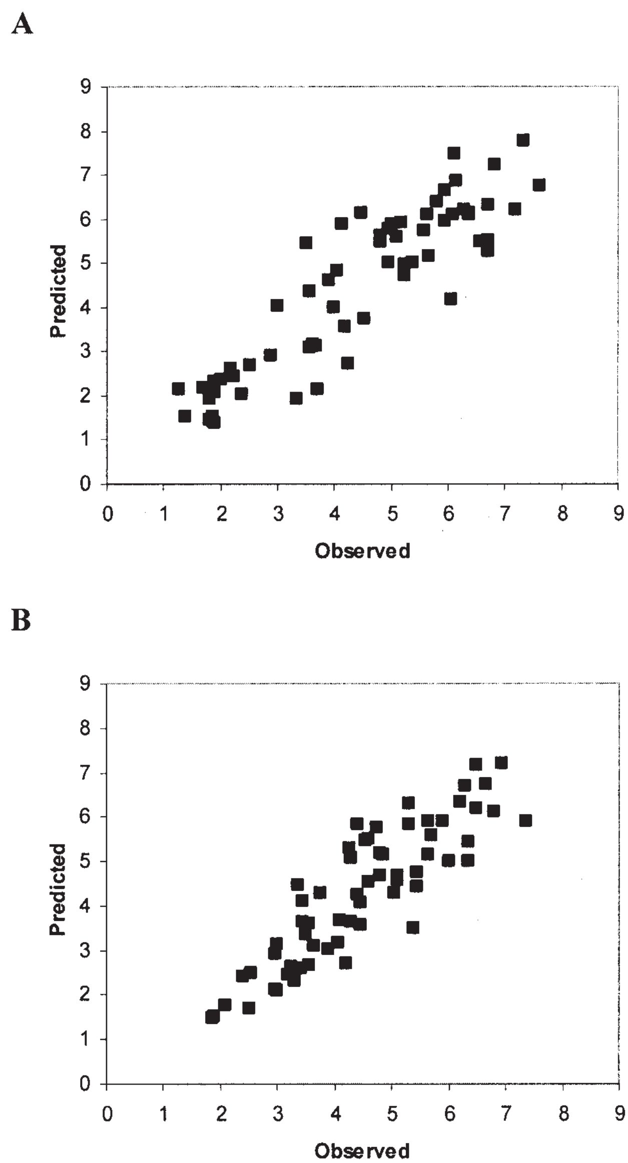

Figure 2. Study 2: Observed and predicted means of typicality ratings for all 63 animals in control participants (A) and in patients with schizophrenia (B).

estimate the typicality of *insect*). Table 2 shows the mean correlations, for averages and standard deviations, between the empirically obtained typicalities and the predicted typicalities based on the error-perturbed data. For comparison purposes, the predictive correlations for the real data reported earlier in this article are also included in this table. Less than 1% of the simulation runs with 20% random instantiations resulted in a better predictive correlation than did the actual patient data. As can be seen in Table 2, the correlations systematically increased as the instantiation-type error decreased, and this was especially evident in the standard deviations.

Second, we introduced randomness at the level of the typicality ratings, the idea being that one aspect of a particular semantic deficit might result in a person rating exemplars' typicality to a superordinate in a fairly random manner. In this second simulation procedure, a fixed percentage of the typicalities of every simulated data set was shuffled, distributed randomly over all instantiations and subordinates. This procedure thus ensured that the overall distribution of the typicality ratings was identical to the distribution of the empirically obtained data. As before, 200 data sets were run for four different percentages. The last columns of Table 2 show the mean correlations, for averages and standard deviations, between the empirically obtained typicalities and the predicted typicalities based on the error-perturbed data. Thus, although the predictive correlations for the averages and standard deviations started off a little higher, the decline as a function of the error percentage was more rapid than for the instantiation-type error.

In summary, the two simulation studies show that the instantiation model is sensitive to randomness in the data. Thus, our study clearly shows that the high predictive correlations for the patient data cannot be dismissed simply by doubting the sensitivity of the model and that these results should be interpreted instead as an indication of the similarity between the patients and the healthy control participants concerning the structure of their semantic knowledge.

General Discussion

In terms of the semantic content of food and animal categories, overall we observed much similarity between patients and control participants. In the category member production task, we did not find appreciable differences between patients and control participants, either in the overall distribution of responses or in rare, idiosyncratic responses. In the typicality rating task, we found some evidence that the group of patients was less consistent than the group of control participants. Given the numerous cognitive components involved in order to perform the rating task and the absence of significant group differences in the production task, we do not regard this as evidence of a group difference of the magnitude or type that could be indicative of an idiosyncratic (or consistent) disorder in category content in patients. Such an isolated finding (in the context of all the other nonsignificant findings)

Table 2
Correlations From Empirical Data, Random Instantiations, and Random Typicality Ratings

	Empirical data		% random instantiations				% random typicality ratings			
Correlation	Control participants	Patients	20	25	50	75	20	25	50	75
M	.90	.88	.86	.86	.82	.61	.88	.88	.65	.34
SD	.74	.70	.64	.64	.52	.32	.72	.71	.25	.08

makes for inherent difficulty in pinpointing a specific cognitive function of a task that is fairly complex. Indeed, data collection in patients with schizophrenia generally results in data with more variability, but how one chooses to interpret such data is of major concern in neuropsychological investigations of patients. Naturally, it could be interesting to further study these deviations specifically to pinpoint the cause of such variance in patients, but first it would be necessary to establish whether there is systematicity in these deviations (i.e., they are consistent). If these deviations are not stable over time (within patients), then it is very likely that the variance does not reflect systematic semantic differences but is more likely indicative of random behavior triggered by environmental or task cues (see Elvevåg & Storms, 2003, and Storms et al., 2003, for further discussion). However, the key result we would emphasize is that the differences between the patient group and the control group were not greater than the differences within the patient group itself.

In terms of structure, there were two key findings. First, our results replicate the findings of Heit and Barsalou (1996) and of De Wilde et al. (2003) that the instantiation model accounts well for the data from healthy control participants. Second, the model is about equally suited for predicting the data from the patients with schizophrenia. The structural relations between superordinate categories and their instantiations, as assumed by the instantiation model, seem to apply to both control participants and patients. Put another way, regardless of any differences in content of beliefs about foods and animals (and there was not much evidence of differences in content), the patients' beliefs seemed to be organized in a systematic way that paralleled the organization of the control participants' beliefs. Although our data revealed little wrong with patients' concepts, it is of course possible that there could be a specific problem in combining concepts into more complex semantic expressions or in specific types of concepts (e.g., emotionally laden concepts). Empirical examination of such a possibility would indeed be interesting.

Our current findings would seem to be at odds with previous findings (e.g., MacKay et al., 1996; Tamlyn et al., 1992). However, we note that unlike previous research, our own tests did not involve speeded judgments or in any other way measure fluency or response time. Likewise, our own tests did not require participants to distinguish factual information from nonfactual information. The production task asked participants to name, at their own pace, a small number of category members, and the typicality rating task asked participants to assess the centrality of category members on the basis of their own beliefs. Of course, it is possible that the problem might not be within the semantic system itself but in its control mechanisms. However, we suggest that this is not a very strong candidate in light of some recent data that we have collected in patients with schizophrenia and in healthy control participants. In that study (Cohen, Elvevåg, & Goldberg, in press), we designed a paradigm in which participants were to determine rapidly whether one of two stimuli was larger or smaller in the real world than the other. Stimuli pairs were either words or their corresponding real-world entities (e.g., *ant* and *house*). We manipulated both real-world "distance" in terms of size (e.g., *ant* and *house* or *ant* and *key*) and the "congruency" between real-world size and physical size on the computer monitor of the image pairs (e.g., *small ant* and *large house* or *large ant* and *small house*), the latter being a task requiring a degree of cognitive control. Crucially, both the distance and congruency effects were qualitatively equivalent across the groups. This suggests that some aspects of semantic knowledge are represented equivalently in patients and control participants and that patients can use this information to override incongruent information in a manner similar to that of control participants (i.e., this aspect of cognitive control is intact).

These findings are consistent with the current data in that the patients had similar representations of categories (see also Elvevåg, Weickert, et al., 2002). In a similar vein, Kéri and colleagues have reported that prototype learning of dot-pattern category exemplars is intact in patients with schizophrenia (Kéri, Kelemen, Benedek, & Janka, 2001) and that patients are able to establish representations of complex categories (Kéri et al., 2000; but see Kéri's other work for different conclusions, namely, that abstraction is impaired in patients, Kéri et al., 1999), although when there is a verbal category definition, patients perform similarly to control participants (Kéri et al., 1998).

In a more general sense, recent data (Elvevåg et al., 2003), suggested that patients with schizophrenia organize information in memory in a qualitatively similar manner, a finding again consistent with the results from the current study showing that the structure of categories is similar in patients and control participants. In this recent study (Elvevåg et al., 2003), patients displayed a qualitatively similar distribution of memories retrieved across the life span in a free-recall (i.e., noncued) autobiographical memory study. Because people acquire representations over a lifetime, if there is a problem in the brain areas (i.e., in the left temporal parietal cortex) that are responsible for this representation process, then one might predict that patients would display a difference in the content or structure of category information (given the neurodevelopmental nature of schizophrenia). Of importance, in the current study we did not find evidence of a difference in either the content or the structure of natural categories in schizophrenia. It is possible, of course, that the integrity of category structure is related to the severity of the deficit. Indeed, in an examination of the consistency of face-naming deficits in patients with schizophrenia across time, researchers concluded that their naming performance varied according to quantitative differences in deficit severity (Laws, McKenna, & Kondel, 1998). Thus, with a more severe patient group our current findings may have been different, although this is probably unlikely given that our patients were fairly chronic. Naturally, it is possible that antipsychotic medication may have interacted with our current findings, because all the patients in our current sample were medicated. Although we suggest that it is unlikely that our findings are simply artifacts of medication, the only way to resolve this issue is to examine it empirically (i.e., in unmedicated patients).

In light of our several studies that have found patients' performance to be indicative of various intact cognitive processes (within the semantic domain), the question emerges as to what could account for patients' "abnormal" behavior. Given the very nature of experimentally fractionating the various processes into their fundamental components, studies become necessarily and increasingly specific as to what subprocesses are intact and compromised. With reference to the clinical manifestation of the illness, such as the deviations in free speech, it is noteworthy that in such experiments there were considerably less constraints than in most experimental language and semantic tasks (including those used in the current study), and thus, there is more scope for unusual associations and problems in the organization of the flow of speech. Indeed, the demanding nature of the dynamic cognitive processing required when assembling all the numerous components involved in free speech (e.g., comprehending, planning, organizing, and composing coherent speech) is disproportionately more difficult (and affected). Thus, in patients with attentional and working memory problems, as well as some positive symptoms, there may be a multiplicative effect that may thus appear as a problem within the semantic domain, at least when evaluated through speech. Nonetheless, on the basis of current findings, it would seem improbable that the observations of patients' unusual content and structure in language reflects a disturbance in their actual knowledge contained in semantic concepts (and the words that are used to refer to these concepts). Fractionating the components of such language processing that *in combination* produce the abnormal behavior characteristic of patients with schizophrenia, especially those with positive symptoms, presents an exciting and useful challenge for future research.

References

Allen, H. A., Liddle, P. F., & Frith, C. D. (1993). Negative features, retrieval processes and verbal fluency in schizophrenia. *British Journal of Psychiatry, 163,* 769–775.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed). Washington, DC: Author.

Cattell, R. B. (1971). *Abilities: Their structure, growth and action.* Boston: Houghton Mifflin.

Cohen, J. R., Elvevåg, B., & Goldberg, T. E. (in press). Cognitive control and semantics in schizophrenia: An integrated approach. *American Journal of Psychiatry.*

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8,* 240–247.

De Wilde, E., Vanoverberghe, V., Storms, G., & De Boeck, P. (2003). The instantiation principle re-evaluated. *Memory, 11,* 533–548.

Elvevåg, B., Fisher, J. E., Gurd, J. M., & Goldberg, T. E. (2002). Semantic clustering in verbal fluency: Schizophrenic patients versus control participants. *Psychological Medicine, 32,* 909–917.

Elvevåg, B., Kerbs, K. M., Malley, J. D., Seeley, E., Weinberger, D. R., & Goldberg, T. E. (2003). Autobiographical memory in schizophrenia: An examination of the distribution of memories. *Neuropsychology, 17,* 402–409.

Elvevåg, B., & Storms, G. (2003). Scaling and clustering in the study of semantic disruptions in patients with schizophrenia: A re-evaluation. *Schizophrenia Research, 63,* 237–246.

Elvevåg, B., Weickert, T., Wechsler, M., Coppola, R., Weinberger, D. R., & Goldberg, T. E. (2002). An investigation of the integrity of semantic boundaries in schizophrenia. *Schizophrenia Research, 53,* 187–198.

Elvevåg, B., Weinstock, D. M., Akil, M., Kleinman, J. E., & Goldberg, T. E. (2001). A comparison of verbal fluency tasks in schizophrenic patients and normal controls. *Schizophrenia Research, 51,* 119–126.

Goldberg, T. E., Aloia, M. S., Gourovitch, M. L., Missar, D., Pickar, D., & Weinberger, D. R. (1998). Cognitive substrates of thought disorder. I: The semantic system. *American Journal of Psychiatry, 155,* 1671–1676.

Goldberg, T. E., Dodge, M., Aloia, M., Egan, M. F., & Weinberger, D. R. (2000). Effects of neuroleptic medications on speech disorganization in schizophrenia: Biasing associative networks towards meaning. *Psychological Medicine, 30,* 1123–1130.

Goldberg, T. E., Torrey, E. F., Gold, J. M., Bigelow, L. B., Ragland, R. D., Taylor, E., & Weinberger, D. R. (1995). Genetic risk of neuropsychological impairment in schizophrenia: A study of monozygotic twins discordant and concordant for the disorder. *Schizophrenia Research, 17,* 77–84.

Heit, E., & Barsalou, L. W. (1996). The instantiation principle in natural categories. *Memory, 4,* 413–451.

Hodges, J. R., Salmon, D. P., & Butters, N. (1992). Semantic memory impairment in Alzheimer's disease: Failure of access or degraded knowledge? *Neuropsychologia, 30,* 301–314.

Institute for Personality and Ability Testing. (1973). *Measuring intelligence with the Culture Fair Test.* Champaign, IL: Author.

Jastak, S., & Wilkinson, G. (1984). *Wide Range Achievement Test—Revised.* Wilmington, DE: Jastak Assessment Systems.

Kaufman, A. (1990). *Assessing adolescent and adult intelligence.* Boston: Allyn & Bacon.

Kéri, S., Kelemen, O., Benedek, G., & Janka, Z. (2001). Intact prototype learning in schizophrenia. *Schizophrenia Research, 52,* 261–264.

Kéri, S., Kelemen, O., Szekeres, G., Bagóczky, N., Erdélyi, R., Antal, A., et al. (2000). Schizophrenics know more than they can tell: Probabilistic classification learning in schizophrenia. *Psychological Medicine, 30,* 149–155.

Kéri, S., Szekeres, G., Kelemen, O., Antal, A., Szendi, I., Kovács, Z., et al. (1998). Abstraction is impaired at the perceptual level in schizophrenic patients. *Neuroscience Letters, 243,* 93–96.

Kéri, S., Szekeres, G., Szendi, I., Antal, A., Kovacs, Z., Janka, Z., & Benedek, G. (1999). Category learning and perceptual categorization in schizophrenia. *Schizophrenia Bulletin, 25,* 593–600.

Kremen, W. S., Seidman, L. J., Faraone, S. V., Pepple, J. R., Lyons, M. J., & Tsuang, M. T. (1996). The "3 Rs" and neuropsychological function in schizophrenia: An empirical test of the matching fallacy. *Neuropsychology, 10,* 22–31.

Laws, K. R., McKenna, P. J., & Kondel, T. K. (1998). On the distinction between access and store disorders in schizophrenia: A question of severity? *Neuropsychologia, 36,* 313–321.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison Wesley.

MacKay, A. P., McKenna, P. J., Bentham, P., Mortimer, A. M., Holbery, A., & Hodges, J. R. (1996). Semantic memory is impaired in schizophrenia. *Biological Psychiatry, 39,* 929–937.

Manschreck, T. C., Maher, B. A., Milavetz, J. J., Ames, D., Weisstein, C. C., & Schneyer, M. L. (1988). Semantic priming in thought disordered schizophrenic patients. *Schizophrenia Research, 1,* 61–66.

Missar, C. D., Gold, J. M., & Goldberg, T. E. (1994). WAIS–R short forms in chronic schizophrenia. *Schizophrenia Research, 12,* 247–250.

Ober, B. A., Vinogradov, S., & Shenaut, G. K. (1995). Semantic priming of category relations in schizophrenia. *Neuropsychology, 9,* 220–228.

Spitzer, M. (1997). A cognitive neuroscience view of schizophrenic thought disorder. *Schizophrenia Bulletin, 23,* 29–50.

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, M., Weisker, I., Winter, M., Maier, S., Hermle, L., & Maher, B. A. (1994). Semantic and phonological priming in schizophrenia. *Journal of Abnormal Psychology, 103,* 485–494.

Storms, G., De Boeck, P., & Ruts, W. (2000). Prototype and exemplar-based information in natural language categories. *Journal of Memory and Language, 42,* 51–73.

Storms, G., Dirikx, T., Saerens, J., Verstraeten, S., & De Deyn, P. P. (2003). On the use of scaling and clustering in the study of semantic deficits. *Neuropsychology, 17,* 289–301.

Sumiyoshi, C., Matsui, M., Sumiyoshi, T., Yamashita, I., Sumiyoshi, S., & Kurachi, M. (2001). Semantic structure in schizophrenia as assessed by the category fluency test: Effect of verbal intelligence and age of onset. *Psychiatry Research, 105,* 187–199.

Tamlyn, D., McKenna, P. J., Mortimer, A. M., Lund, C. E., Hammond, S., & Baddeley, A. D. (1992). Memory impairment in schizophrenia: Its extent, affiliations and neuropsychological character. *Psychological Medicine, 22,* 101–115.

Vinogradov, S., Ober, B. A., & Shenaut, G. K. (1992). Semantic priming of word pronunciation and lexical decision in schizophrenia. *Schizophrenia Research, 8,* 171–181.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised.* San Antonio, TX: Psychological Corporation.

Weickert, T. W., Goldberg, T. E., Gold, J. M., Bigelow, L. B., Egan, M. F., & Weinberger, D. R. (2000). Cognitive impairments in patients with schizophrenia displaying preserved and compromised intellect. *Archives in General Psychiatry, 57,* 907–913.

Wiens, A., Bryan, J., & Crossen, J. (1993). Estimating WAIS–R FSIQ from the National Adult Reading Test—Revised in normal subjects. *The Clinical Neuropsychologist, 7,* 70–84.

Received May 13, 2003
Revision received May 24, 2004
Accepted May 25, 2004 ■