Neuropsychological Trends

2
November 2007

First-episode schizophrenia: characterization and clinical correlates

Robert M. Bilder - Robert S. Goldman
Delbert Robinson - Gail Reiter - Lisa Bell - John A. Bates
Elizabeth Pappadopulos - Jose Maria J. Alvir
Margaret G. Woerner - Stephen Geisler - John M. Kane

David Geffen School of Medicine and Psychology, UCLA

mbilder@mednet.ucla.edu

Abstract

Neuropsychological impairments are well documented in schizophrenia and are important targets of treatment. Information about the severity and pattern of deficits after treatment for the first psychotic episode and about relationships between these deficits and syndromal characteristics remains limited. Comprehensive neuropsychological assessments including 41 individual tests were given to 94 patients with first-episode schizophrenia after initial stabilization of psychosis and to a comparison group of 36 healthy volunteers. Profiles of neuropsychological deficits and the relationship of deficits to sex and handedness were examined. Correlations of neuropsychological deficit with a broad range of historical and clinical characteristics, including outcome, were explored. Patients had a large generalized neuropsychological deficit. Patients also had, superimposed on the generalized deficit, subtle relative deficits in memory and executive functions. Learning/memory dysfunction best distinguished patients from healthy individuals; after accounting for this difference, only motor deficits further distinguished the groups. Patients with higher neuropsychological ability had only memory deficits, and patients with lower ability had both memory and executive deficits. Dextral patients had less severe generalized deficit. Severity of residual symptoms was associated with greater generalized deficit. Executive and attentional deficits were most linked to global functional impairment and poor outcome. The results document a large generalized deficit, and more subtle differential deficits, in clinically stabilized first-episode patients. Learning/memory deficits were observed even in patients with less severe generalized deficit, but the pattern was unlike the amnestic syndrome and probably reflects different mechanisms. Executive and attentional deficits marked the more severe-

ly disabled patients, and may portend relatively poor outcome. Failure to develop typical patterns of cerebral dominance may increase the risk for greater generalized deficit.

Keywords: Neuropsychology; Schizophrenia Spectrum Disorders; Learning; Memory

1. Introduction

Neuropsychological deficits are recognized as an important pathologic dimension in schizophrenia. These deficits are often severe and pervasive, but estimates of severity have varied, knowledge about when deficits first appear is incomplete, and the interpretation of deficit patterns remains controversial. Some researchers have emphasized the generalized deficit and highlighted the psychometric hurdles that must be overcome to identify differential deficits (Heaton et al., 1978; Goldstein, 1978; Blanchard & Neale, 1994; Mohamed, 1999). Others have emphasized relative deficits on tasks that are putatively sensitive to dysfunction of specific brain systems, including (1) the frontal lobes (Goldberg et al., 1987), (2) the temporal and/or mesiotemporal regions (Saykin et al., 1991; 1994), and (3) integrated frontotemporal or frontolimbic systems (Bilder et al., 1995; Bilder & Degreef, 1991; Bilder & Szeszko, 1996; Flor-Henry, 1983; Gruzelier et al., 1988; Gruzelier, 1991). Although attempts to localize brain dysfunction in schizophrenia on the basis of neuropsychological tests have had multiple problems (Bilder & Goldberg, 1987), specific deficit patterns may nevertheless pose significant constraints on models of pathophysiology. Furthermore, substantial interest is currently focused on the capacity of novel treatments to ameliorate neuropsychological deficits. Given suggestions that neuropsychological deficits are linked to disability (Green, 1996), treatments that are beneficial for cognition may also attenuate other functional limitations. Most studies examining neuropsychological deficits in schizophrenia have involved chronic patients, and the findings may reflect the long-term illness or treatment experienced by these patients. These study groups may also be biased toward poorer treatment response and outcome. Some results from groups of first-episode patients have been reported (Mohamed et al., 1999; Saykin, 1994; Bilder et al., 1992; Hoff et al., 1992; DeLisi et al., 1995; Nuechterlein et al., 1995; Censits et al., 1997).

These studies have shown that deficits are present at illness onset, but their conclusions have been limited by a variety of factors, such as small size of the study group, lack of standardized treatments, patients' lack of clinical stability at the time of assessment, or lack of longitudinal clinical follow-up. This report provides comprehensive neuropsychological characterization of patients after the first episode of illness, when most patients satisfied criteria for clinical stabilization after standardized treatment with conventional antipsychotics. The strategy of assessing first-episode patients after clinical stabilization aimed to maximize measurement of stable trait-like neuropsychological characteristics and to minimize transient effects associated with acute psychosis and its treatment. The results of a longitudinal neuropsychological follow-up over 5 years and of neuropsychological tests administered before treatment, as well as descriptions of correlations with anatomic measures from magnetic resonance imaging, are being reported separately.

We previously reported preliminary neuropsychological findings in a subgroup of the patients described here (Bilder et al., 1995; 1992; 1991; Szeszko et al., 1999) and recently published papers about treatment response and outcome in the overall group of 118 patients from which the patients described here were selected (Robinson et al., 1999a; 1999b). The study reported here examined (1) the magnitude of neuropsychological deficit in patients with schizophrenia after clinical stabilization of the first episode of psychosis, (2) the pattern of neuropsychological deficits in this group, and (3) the nature of any neuropsychological differences between groups defined by sex or handedness. We further examined relations of neuropsychological deficits with a broad range of historical, symptomatic, treatment response, and outcome characteristics to provide descriptive information about the possible selectivity and magnitude of these associations.

2. Methodology

2.1. Subjects

Subjects were participants in the Prospective Study of Psychobiology in First-Episode Schizophrenia at Hillside Hospital in Glen Oaks, N.Y. (Robinson et al., 1999a; 1999b; Lieberman et al., 1992; 1993). Patients who were admitted to the hospital inpatient service for a first episode of psychotic illness and who had received less than 12 prior weeks of cumulative lifetime neuroleptic treatment were recruited for the study. Patients satisfied Research Diagnostic Criteria (RDC) (Spitzer et al., 1977; 1978) for schizophrenia or schizoaffective disorder, on the basis of structured interviews with the Schedule for Affective Disorders and Schizophrenia (SADS) (Endicott & Spitzer,

1978) and reviews of patients' histories. Patients with a current or past serious neurologic or endocrine disorder were excluded.

After a complete description of the study was provided to the subjects, we obtained their written informed consent. Further details of ascertainment and treatment have been published elsewhere (Robinson et al., 1999a; 1999b; Lieberman et al., 1992; 1993). The healthy comparison group was recruited through announcements in local newspapers and within the medical centre. These subjects were selected to be similar to the patients on distributions of sex and age. They were free of RDC mental disorders, as determined by using the SADS Lifetime Version interview, physical examination, and urinalysis (Lieberman et al., 1992). None of the subjects had a current substance use disorder or a history of substance dependence, chronic neurologic or medical illness, or drug treatment known to affect the brain.

2.2. Procedure

Neuropsychological tests were planned for 6 months after study entry if the patient had already achieved remission or a stable level of residual symptoms. We selected 6 months because pilot data showed asymptotic levels of symptom remission at that point. Patients who did not satisfy the criteria for symptom remission or stable residual symptoms at 6 months were tested as soon as possible after they satisfied the criteria. Remission was defined as (1) a rating not greater than 3 on the positive psychotic symptoms items of the SADS Change Version (SADS-C) psychosis and disorganization dimensions, (2) a rating of 3 (mild) or less on the Clinical Global Impression (CGI) severity items, (3) a rating of 2 (much improved) or 1 (very much improved) on the CGI improvement items, and (4) maintenance of this level of response for 8 weeks. Stability of residual symptoms required no changes greater than 1 point on positive psychotic symptom items of the SADS-C psychosis and disorganization dimensions or on global ratings on the Scale for the Assessment of Negative Symptoms (SANS) for two consecutive biweekly rating periods.

Test administration was typically divided into two 3-4 hour sessions within 1 week, and two counterbalanced test sequences were employed across subjects to minimize possible order effects. The battery included 41 tests from which variables were selected to characterize six neuropsychological domains represented by the following scales: language, memory, attentional, executive, motor, and visuospatial. A global neuropsychological scale that represented the mean of these six scales was constructed. A scale for measuring premorbid intellectual ability was also constructed, based on the idea that certain tests of general knowledge, vocabulary, and reading skill are less vulnerable to deterioration, as our previous reports have shown (Bilder et al., 1992; 1988).

Loading of test variables on scales was based on a priori assessment of content validity, by using methods similar to those used previously by our group (Bilder et al., 1995; 1988; Bilder, 1985; Bilder et al., 1985) and others (Blanchard et al., 1994; Saykin et al., 1991; 1994; Hoff et al., 1992). Scores for each scale were computed by averaging z scores on contributing variables. These z scores were based on performance of the healthy comparison group, which by definition had mean scale scores of zero and standard deviations set to one. All scales were computed so that higher values indicated better performance. At each stage of scale construction, contributing variables were restandardized before means were computed over all nonmissing data. The values for widely used test variables were not restandardized; the original scores, t tests for the difference between groups, and the 95% confidence intervals (CIs) for the difference in original score units, are provided for descriptive purposes.

Additional scaling procedures were applied to improve psychometric properties, and confirmatory factor analysis was used to assess validity of the assignments of variables to scales. First, each test variable was examined for extreme values, and in several instances these deviant scores were replaced by scores within the tails of their underlying distributions (this procedure affected two tests and a total of seven patients' scores). Second, we examined each variable for possible sex differences. Only the Finger Tapping Test showed a significant difference, with men having higher scores, and therefore variables from this test were standardized separately in each sex. Third, the distributions were examined both within and between groups, with special attention to problems involving heteroscedasticity, and variance-stabilizing algorithms were applied (in most cases power transformations) to optimize homogeneity of variance between groups for each variable (Levene's test was used as a criterion). Fourth, reliability analyses were conducted for each scale, using the initial a priori variable assignments, and any test variable that decreased the internal consistency of the composite (as assessed by Cronbach's coefficient alpha) was eliminated. Fifth, the scales were computed, and at each stage of scale construction, further tests on homogeneity of variance were conducted and transformations were applied where indicated, as was done for the individual test variables. Finally, the validity of variables' assignments to scales, which was solely on a rational basis, was examined using confirmatory factor analysis (Bentler & Bonnett, 1980).

This analysis supported the rational grouping of neuropsychological variables into the scales described here (comparative fit index = 0.92 [2 = 1593, df = 29, $p < 0.001$] for the comparison with the null model; comparative fit index = 0.86 [2 = 1494, df = 24, p < 0.001] for the comparison with the single-factor model). Scales with higher reliability are more sensitive to group differences than less reliable scales. In this study the standard error of measurement varied from 0.21 to 0.46 across scales, offering relatively high uniformity and precision in the estimate of true scores compared to other reports in the literature. Although alternate models might yield superior results, systematic exploration of these models is beyond the scope of this paper.

Hand preference was assessed using a modified 20-item Edinburgh Inventory (Oldfield, 1971). The total number of right- and left-hand items were scored, and the laterality quotient was computed as (total right-total left) / (total right + total left). Thus laterality quotients could range from 1.00 (all items right) to -1.00 (all items left). Subjects with a laterality quotient greater than 0.70 were classified as dextral; the rest were classified as nondextral (Schachter et al., 1987).

3. Data analysis

The primary tests of group differences used profile analysis by multivariate analysis of variance (MANOVA), with group as a between-subjects factor and neuropsychological scale as a within-subject factor. Multivariate analysis of covariance (MANCOVA) was used to assess effects of possible moderating variables. Deviations from flatness in the patient profiles, if suggested by significant effects of the interactions of group and scale, were assessed by contrasting the mean for each individual scale with the mean of all other scales using paired t tests. This procedure allowed us to determine which scales showed impairment relative to each of the other scales in the analysis of data from the patient group. We also used the method recommended by Chapman and Chapman (1986), involving standardized residual scores. In contrast to the profile analysis by MANOVA (which examines the extent to which the average score on Scale X differs from the mean score on other scales), standardized residual score analysis considers the statistical interdependence of the scales (i.e, examines the extent to which a subject's score on Scale X is deviant, given that subject's scores on the other scales). Prediction equations were based on data from the comparison group.

The 94 patients completing neuropsychological exams satisfied RDC for schizophrenia (N = 70; subtypes included: paranoid = 54, disorganized = 4, catatonic = 1, and undifferentiated = 11) or schizoaffective disorder (N = 24). Characteristics of the patient group and the comparison group are provided in Table 1.

Table 1. Characteristics of patients with Schizophrenia and healthy comparison subjects in a study of neuropsychological impairments after treatment for first-episode psychosis

Characteristics	Patient group		Comparison group		Analisys		
					Test statistic	*df*	*p*
	N	%	*N*	%			
Male sex	55	58.5	24	66.7	$X^2 = 0.73$	1	n.s.
Race					$X^2 = 15.90$	4	0.01
White	42	44.7	28	77.8			
African American	34	36.2	2	5.6			
Ispanic	10	10.6	2	5.6			
Asian	6	6.4	2	5.6			
Other	2	2.1	2	5.6			
Right-Handedness	67	71.3	27	75.0	$X^2 = 0.18$	1	n.s.
	Mean	*SD*	*Mean*	*SD*			
Age (years)	25.7	6.3	26.3	6.5	T = 0.28	128	n.s.
Parental Social Class[a, b]	3.1	2.6	2.6	1.1	T = 2.23	123	0.03
Education (years)	13.1	2.3	16.9	1.5	T = 4.55	128	0.001
Laterality quotient[a]	0.71	0.49	0.73	0.43	T = -0.37	127	n.s.

[a] Data on parental social class were missing for one patient and four comparison subjects; data on laterality were missing for one patient.

[b] Holingshead-Redich system, in which 1 = highest and 6 = lowest.

The groups were well matched on distributions of sex, age, and hand preference, but they differed in racial/ethnic group composition, parental social class, and education. Possible effects of these differences were examined in subsequent analyses.

The 94 patients, who completed at least one comprehensive neuropsychological examination, were a subset of the 118 patients described by Robinson et al. (1999a; 1999b), reflecting an 80% completion rate. To determine if the 94 patients were representative of the larger group, we compared the 94 patients who completed the neuropsychological examinations with the 24 patients who did not. The two groups were similar in age, education, racial/ethnic group composition, parental social class, RDC diagnosis, duration of symptoms before study entry, measures of symptoms at baseline, and course/outcome characteristics (all p values > 0.05). Compared to the patients who completed the neuropsychological examinations, those who did not were more likely to be women (75%, N = 18, compared to 41%, N = 39; 2 = 8.6, df = 1, p = 0.003) and to have been married (42%, N = 10, compared to 16%, N = 15; 2 = 24.0, df = 3, p = 0.00002). Among the 94 patients, nine patients did not satisfy stabilization criteria at 6 months when they were initially scheduled for examination. They were reexamined during the next year, after they had achieved the criteria. Twelve patients who entered the study before the neuropsychological protocol was put in place (April 1988) were examined as soon as possible after the protocol was established. The modal time from the beginning of treatment to neuropsychological examination was 0.47 years (median = 0.61 years); 79% (N = 74) of the patients were seen within the first year of any treatment.

The patient group was more impaired than the comparison group on every neuropsychological dimension measured. Mean effect sizes (in z score units, reflecting the number of standard deviations below the comparison group means) ranged from -1.11 to -1.75; 95% CIs ranged from a minimum deficit of -0.82 to a maximum deficit of -2.13 (main effect of group: F = 87.2, df = 1, 123, p < 0.0001, N = 125). The overall profile mean for the patients was -1.53, indicating a generalized deficit of approximately 1.5 standard deviations.

The patient profile deviated significantly from flatness; in other words, the means for some scales reflected more impairment than the means for other scales (group-by-scale interaction: Wilks lambda = 0.91; F = 2.39, df = 5, 119, p < 0.04, N = 125). Specifically, the memory (paired t = 3.44, df = 88, p = 0.001, N = 90) and executive (paired t = -2.36, df = 88, p = 0.02, N = 90) scales showed significantly more impairment, and the language scale (paired t = 5.64, df = 88, p < 0.001, N = 90) showed significantly less impairment, compared with the remaining scales. Additional analyses examined possible effects of parental social class (by using MANCOVA), racial/ethnic group composition (by examining results for white subjects, the only racial/ethnic group large enough for analysis), and diagnostic subgroups (by examining results for patients with RDC schizophrenia only compared to the healthy subjects). None of these analyses produced findings that differed substantively from those of the original analyses. No significant differences in neuropsychological profiles were found between the patients with schizophrenia and those with schizoaffective disorder.

Analysis of standardized residual scores revealed a significant group effect (Wilks lambda = 0.77, F = 5.84, df = 6, 118, p = 0.001, N = 125), with patients showing significant residual deficits on memory (F = 5.70, df = 1, 123, p = 0.02, N = 125) and motor (F = 6.65, df = 1, 123, p = 0.01, N = 125) scales. Because the generalized deficit of the patients was so large, we examined differences in neuropsychological profiles between groups of patients with low and high levels of general ability (median split on the global neuropsychological scale). The high-ability group (N = 43) had a mean deficit of -0.83 standard deviations (95% CI = -0.98 to -0.70), and the low-ability group (N = 46) had a mean deficit of -2.22 standard deviations (95% CI = -2.36 to -2.07), relative to the comparison group. MANOVA showed that these groups differed in profile (Wilks lambda = 0.65; F = 9.03, df = 5, 83, p < 0.001).

The high-ability group showed relative deficits (compared to their own profile mean) only on the memory scale (mean deficit = -1.07, 95% CI = -1.26 to -0.87). The low-ability group showed relative deficits on both the memory scale (mean deficit = -2.48, 95% CI = -2.68 to -2.28) and the executive scale (mean deficit = -2.50, 95% CI = -2.75 to -2.25). The low-ability group also showed a higher mean language scale score than would be expected based on their profile mean (mean deficit = -1.60, 95% CI = -1.82 to -1.38). Effects of sex were assessed using MANOVA with diagnostic group (patient group or comparison group) and sex as between-subject factors and the six neuropsychological scales as within-subject factors. The main effect of sex (F = 0.69, df = 1, 121, p = 0.40, N = 125), and all of the interactions involving sex were nonsignificant (all F values < 1). The main effect of group and the effect of the interaction of group and scale were essentially identical to the original analysis. In addition, we examined the sex difference on the scale that measured premorbid functioning.

Significantly greater impairment was found in the female patients compared to the male patients (t = 2.89, df = 92, p < 0.005, N = 94); no sex differences were observed in the comparison group. Handedness effects could not be assessed using MANOVA due to a violation in the assumption of multivariate homogeneity of variance (Box's M = 121; approximate F = 1.61, df = 63, 3297, p < 0.002). We therefore examined differences using t tests. No significant effects of handedness were found in the comparison group. In contrast, the 67 dextral patients performed approximately 0.5 standard deviations higher on each scale compared to the 27 nondextral patients. These groups differed in global neuropsychological scale score by 0.47 scaled score units (mean deficit for dextral patients = -1.65, SD = 0.87; mean deficit for nondextral patients = -2.12, SD = 0.84; t = 2.4, df = 92, p = 0.02). Differences were also observed on the scales for the neuropsychological domains (range of differences on individual scales = 0.35-0.63, range of t values = 1.6-2.6, df = 92, range of p values = 0.11-0.01). No significant differences in the distribution of handedness were found among male and female patients. To examine the possibility that a subgroup of patients with mixed hand preference might be particularly impaired (Green et al., 1989), we inspected scatterplots of neuropsychological scale scores with respect to the laterality quotient.

Table 2. Correlation between clinical variables and neuropsychological scale scores among 94 patients with first-episode schizophrenia

Variable	Correlation with Neuropsychological Scale Score (r, two tailed)							
	Language	Memory	Attention	Executive	Motor	Visuospatial	Premorbid functioning	Global
Symptoms at times of neuropsychological examination								
Schedule for affective disorder and Schizophrenia, psychosis and disorganization dimensions								
Delusions	-0.12	-0.26	-0.24	-0.27	-0.19	-0.19	-0.12	-0.27
Hallucinations	-0.11	-0.15	-0.12	-0.29	-0.17	-0.19	-0.12	-0.22
Thought Disorder	0.12	-0.18	-0.15	-0.12	-0.13	-0.06	0.02	-0.08
Bizarre Behaviour	-0.01	-0.11	-0.08	-0.08	-0.15	-0.08	-0.02	-0.10
Scale for the Assessment of Negative Symptoms								
Affective Flattering	-0.21	-0.27	-0.30*	-0.39**	-0.27	-0.34	-0.16	-0.38**
Alogia	-0.16	-0.34**	-0.34**	-0.41**	-0.23	-0.32*	-0.16	-0.37**
Apathy	-0.09	-0.08	-0.28	-0.30*	-0.34**	0.35**	0.00	-0.31*
Anhedonia	-0.05	-0.22	-0.23	-0.28	-0.33*	-0.23	-0.05	-0.28
Attention Impairment	-0.19	-0.30	-0.27	-0.30*	-0.19	-0.13	-0.18	-0.26

Global functioning and outcome indices								
Childhood Premorbid Adjustment Scale	-0.24	-0.27	-0.31*	-0.22	-0.21	-0.11	-0.23	-0.27
Global Assessment Scale	0.18	0.34**	0.37**	0.53**	0.38**	0.45**	0.11	0.47**
Course (first episode)	-0.14	-0.31	-0.25	-0.31*	-0.21	-0.28	-0.11	-0.29
Social Adjustment Scale, general adjustment (year 2)	-0.07	-0.23	-0.27	-0.19	-0.22	-0.03	-0.21	-0.21
Treatment measures								
Antipsychotic dose at testing[a]	-0.28	-0.36**	-0.30*	-0.40**	-0.27	-0.27	-0.23	-0.38**
Cumulative antipsychotic dose[a]	-0.15	-0.23	-0.09	-0.12	-0.04	-0.14	-0.11	-0.13
Antiparkinson dose[b]	0.0	-0.35*	-0.17	-0.26	-0.20	-0.17	0.00	-0.27

[a] Chlorpromazine equivalents (N = 85).
[b] Benztropine equivalents (N = 72).
* $p < 0.05$.
** $p < 0.001$.

We then examined the effects of adding a quadratic term to the linear relation between the laterality quotient and the global neuropsychological scale score within the patient group. The overall regression analysis was significant ($F = 6.2$, $df = 2, 90$, $p < 0.003$, $N = 94$), but the significance was mostly due to the linear term (beta = 0.41, $T = 3.5$, $p < 0.007$), supporting the original analysis. The quadratic term contributed less (beta = -0.18, $T = -1.5$, $p = 0.13$), and the sign of this term was opposite that predicted if patients with mixed handedness performed more poorly. There were no significant correlations of laterality quotient with neuropsychological scale scores in the comparison group.

We explored relationships between neuropsychological scale scores and demographic, historical, and clinical variables, including:

1. age at time of examination,
2. age at onset of first psychotic symptoms,
3. ratings on the NIMH modification of the Premorbid Adjustment Scale (Cannon-Spoor et al., 1982),
4. global ratings of extrapyramidal symptoms made after 8 and 16 weeks of treatment,
5. global measures of course made by physicians after patients completed the first year of the study,
6. ratings on the Social Adjustment Scale made after 2 years in the study,
7. ratings of the deficit syndrome (Kirkpatrick et al., 1989),
8. medications prescribed at the time of testing. Further descriptions and operational criteria are published elsewhere (Bilder et al., 1992; Robinson et al., 1999a; 1999b; Lieberman, 1992; 1993; Mayerhoff et al., 1994; Lieberman et al., 1993).

Because the goal of these analyses was descriptive, and given the large number of tests conducted, only relations with an effect size equivalent to $r > 0.30$ were interpreted as significant. Use of this threshold protected against type I error and helped ensure that reported findings are likely to be replicated. A sample size of 94 yields 85% power to detect an effect of this size with alpha set at 0.05 (two-tailed). The probability of observing $r > 0.30$ in this sample is approximately 0.005 (two-tailed). The standard error of these correlations is approximately 0.10. Thus we interpreted only effects with probability less than 0.005 (two-tailed), and we can be 95% confident that the true population correlations were within 0.2 of those reported. Some of these results are described below, and others are shown in Table 2.

There were significant correlations between lower neuropsychological scores and poorer Premorbid Adjustment Scale ratings, which were most consistent for social-personal adjustment (e.g., with global neuropsychological scale score, $r = -0.34$, $df = 88$, $p = 0.0009$) but were not specific to any single neuropsychological function. There were moderate relations of neuropsychological impairment with extrapyramidal symptoms assessed over the first 16 weeks of treatment; patients who did develop extrapyramidal symptoms (N = 51) had slightly poorer functioning than those who did not develop extrapyramidal symptoms (N = 40) (global neuropsychological scale score: $F = 5.26$, $df = 1, 89$, $p = 0.01$), but did not differ in profile. There were no significant correlations of neuropsychological scales with age in either the patient group or the comparison group, with age at onset in the entire patient sample, or with age of onset among men or women patients.

We examined correlations of scores on each neuropsychological scale with symptom ratings at two time points: (1) at study entry, before treatment (baseline ratings); and (2) close to the time of the neuropsychological examination (typically within 2 weeks of the examination). Ratings on the SADS-C psychosis and disorganization dimensions and on the SANS were examined. There were no correlations greater than 0.30 between neuropsychological scale scores and ratings of symptoms in the SADS-C psychosis and disorganization dimensions at baseline. Correlations with baseline SANS ratings were somewhat more robust; specifically, the SANS global score for affective flattening had correlations > 0.30 with scores on the neuropsychological scales for memory ($r = -0.31$) and attention ($r = -0.36$) and on the global neuropsychological scale ($r = -0.30$) ($df = 91$; all p values < 0.01, two-tailed). Neuropsychological scale scores correlated more strongly with symptom ratings at the time of the neuropsychological examination (table 2). Neuropsychological scale scores tended to correlate most with global clinical assessments and ratings of negative symptoms. Neuropsychological deficits explained approximately 5% to 25% of the variance in ratings of course and general social/vocational outcome after 2 years.

Scores on the executive scale, compared to other neuropsychological scales, appeared to correlate strongly with several functional indices. Given the theoretical implications of differential associations with executive compared to memory deficits, we tested differences between the size of these correlations (Meng et al., 1992). Compared to memory deficits, executive deficits were more strongly correlated with the Global Assessment Scale rating (for the difference between correlations, $t = 2.33$, $df = 91$, $p < 0.01$, N = 94) and the mean of SANS global ratings (for the difference between correlations, $t = 1.66$, $df = 91$, $p < 0.05$, N = 94).

We examined relationships between neuropsychological scores and treatments prescribed and cumulative antipsychotic dose at the time of testing. Doses of antipsychotics were converted into chlorpromazine equivalents, and doses of antiparkinsonian agents were converted into benztropine equivalents. The median chlorpromazine-equivalent dose was 500 mg/day (mean = 712, SD = 730, with a range from 0 mg/day [nine patients were receiving no medication] to 3750 mg/day). The median benztropine-equivalent dose among the 71 patients receiving antiparkinsonian agents was 4 mg/day (mean = 4.56, SD = 2.11, with a range from 0 to 11.5 mg/day). Scores on neuropsychological scales showed moderate correlations with chlorpromazine-equivalent doses (Table 2). (Correlations involving medication doses excluded data for patients who were not receiving medications). However, chlorpromazine-equivalent dose also correlated significantly with a variety of symptom ratings, including severity of hallucinations ($r = 0.55$, $df = 83$, $p < 0.001$) and the CGI score ($r = 0.43$, $df = 83$, $p < 0.001$). After statistically controlling for both of these clinical variables, correlations of chlorpromazine-equivalent dose with neuropsychological scale scores were partially attenuated (e.g., $r = -0.30$, $df = 80$, $p < 0.005$, for the correlation with the memory scale score, and $r = -0.28$, $df = 80$, $p = 0.009$, for the correlation with the global neuropsychological scale score; correlations with scores on other neuropsychological scales ranged from $r = -0.14$ to $r = -0.26$, $df = 80$, p values > 0.02). Benztropine-equivalent dose was significantly correlated with poorer performance on the memory scale, but benztropine-equivalent dose was also significantly correlated with the CGI score ($r = 0.30$, $df = 69$, $p = 0.01$) and had similar correlations with scores on SANS items reflecting decreased spontaneous movement ($r = 0.28$, $df = 69$, $p = 0.02$) and poverty of speech ($r = 0.25$, $df = 69$, $p = 0.04$). After controlling for the CGI scores and scores on these two SANS items, the correlation of benztropine-equivalent dose with memory impairment was attenuated only slightly ($r = -0.25$, $df = 66$, $p < 0.04$).

4. Discussion and conclusion

This study characterized the neuropsychological function of patients after initial stabilization of the first episode of schizophrenia or schizoaffective disorder. The patients showed a generalized deficit of approximately 1.5 standard deviations relative to the comparison group. This effect is large in both statistical and clinical terms. In the context of this generalized deficit, language function was relatively spared, memory was most impaired, and executive and motor dysfunctions also emerged as relative deficits (see below). Although these fluctuations are statistically significant, their magnitude pales in contrast to the size of the generalized deficit, prompting questions about their pathophysiological and clinical significance, as noted recently by Mohamed and colleagues (Mohamed et al., 1999).

Classical neuropsychological interpretation of the mean patient profile suggests a relatively nonspecific deficit pattern, which could reflect either diffuse dysfunction or disturbances in key systems (mesencephalic, diencephalic, limbic, or frontal functional systems) that have modulatory efffects on broadly distributed neural networks. Given the caveat that neuropsychological inference based on adult, focal lesion studies may be invalid in the study of schizophrenia (Bilder et al., 1987; Bilder, 1992), the findings are nevertheless consistent with current conceptualizations of the prevailing neuropsychological deficits in schizophrenia as either "widespread" (i.e., affecting intrinsic cortical circuitry) or as affecting frontolimbic and/or brainstem systems.

The neuropsychological profile observed in this study is similar to results obtained elsewhere in groups of first-episode patients, despite major differences in treatment conditions and more subtle differences in neuropsychological tests and data analytic methods (Mohamed et al., 1999; Saykin et al., 1994; Hoff et al., 1992). The combined results suggest that this profile is a relatively constant feature of the syndrome early in its course. The pattern of deficit is also generally consistent with studies of chronic schizophrenia (Saykin et al., 1991; Heaton et al., 1994), but the overall severity of the deficit is about 0.3 to 1.0 standard deviations greater in the groups of chronic patients compared to first-episode patients (Bilder et al., 1992; Hoff et al., 1998). It remains unclear whether this discrepancy reflects sampling bias (i.e., as many first-episode patients will not go on to have chronic illness) or a deteriorating course and associated cognitive decline in some patients. Recent reports have shown little change and some improvement in first-episode groups followed up to 5 years (Hoff et al., 1999; Gold et al., 1999), but it is difficult to rule out subgroup effects given that only 15%-25% of patients are expected to decline (Bilder, 1992). Further, despite improvement of scores, patients may not show normal gains on retesting, which might reflect functional decline (Hoff et al., 1999).

Some investigators have focused on relative deficits in learning and memory to implicate mesiotemporal pathology (Saykin et al., 1991; Tamlyn et al., 1002), but comparison of memory impairment in schizophrenia to the amnestic syndrome may be misleading. First, memory deficit was not selective enough to resemble amnesia (Squire & Zola-Morgan, 1991); other domains (executive, motor) were equally affected. Second, the discrepancy between immediate and delayed memory was not comparable to that observed in amnesia (Table 1; for detailed analysis of these discrepancies, see Wilson, 1997). Third, memory dysfunction overlapped statistically with other deficits. Memory tests thus appear sensitive to the cognitive pathology of schizophrenia, but probably tap more complex and multifactorial pathology for most patients. Despite these findings across the group, 28% of the patients may satisfy objective criteria for amnesia based on discrepancies between IQ and memory (Goldman et al., 1999), and this subgroup merits further attention. The contrast between patients of high and low neuropsychological ability indicates that learning/memory deficit is present even in patients with a less severe generalized deficit. But in more severely affected individuals, relative deficits in executive functions are also present, and these deficits may be more severe than that for memory. These findings are consistent with the hypothesis of dysfunction in an integrated frontolimbic system, with less severely affected patients showing only learning/memory deficit and more severely affected patients manifesting deficits in executive functions as well (Bilder & Szeszko, 1996).

Motor dysfunction emerged as a relative deficit in analysis of residual scores (i.e., after statistically controlling for all the other scores for that subject). Our finding that motor deficits are statistically independent from memory and other deficits is similar to that of Sullivan and colleagues (Sullivan et al., 1994). Motor deficits may reflect in part adverse medication effects, as seen in studies of acute neuroleptic treatment (Medalia et al., 1988; Bilder et al., 1992). This idea is further supported by previous findings of less severe motor deficits in studies of mostly neuroleptic-naive patients (Mohamed et al., 1999; Saykin et al., 1994). Although an iatrogenic contribution is possible, it is unlikely to be the sole cause of the motor deficits. First, motor dysfunction was neither highly correlated with current or cumulative antipsychotic dose, nor with extrapyramidal symptoms during early treatment. Second, longitudinal analyses in our sample show that motor deficits are present before treatment, are exacerbated acutely by antipsychotic treatment (i.e., over the first few months), and then gradually return to baseline levels with continued treatment (Bilder & Bates, 1994). Third, motor abnormalities have been observed in high-risk samples (Fish et al., 1992; Marcus et al., 1985; Mednick et al., 1974; Neumann et al., 1995) and in home movies of children who later developed schizophrenia (Walker & Levine, 1990).

Although it will be important to examine motor function in patients who receive new antipsychotics with less adverse extrapyramidal effects, the motor impairments reported here likely reflect persistent deficits. Because these motor deficits are statistically independent from other neuropsychological deficits, they may reflect a distinct pathologic process. Correlations of neuropsychological measures with clinical measures had small-to-medium effect sizes; the strongest indicate approximately 25% shared variance. Neuropsychological performance had little relation to symptoms at the time of study entry, but it was correlated with symptoms after clinical stabilization. This finding suggests that symptom assessments of drug-naive patients may offer little insight about persistent deficits. Neuropsychological scales tended to correlate more strongly with negative symptoms than with positive symptoms or symptoms of conceptual disorganization, but the neuropsychological correlations with some positive symptoms were not smaller than the correlations with negative symptoms. This observation suggests that persistent, treatment-resistant positive symptoms may also index a trait-like deficit. The lack of neuropsychological correlations with disorganization symptoms should be considered with caution: this sample had low levels of these symptoms, and language function was relatively well preserved. These observations stand in contrast to findings obtained by using similar methods in a group of chronic patients with both prominent disorganization symptoms and prominent language impairment, where these domains were strongly correlated (Bilder et al., 1985). Patients with these features may be underrepresented in groups of first-episode patients that include more individuals with good outcome, or these features may emerge later in the illness.

Neuropsychological measures correlated significantly with both childhood adjustment and current global functioning indices, suggesting an early developmental origin of enduring deficits. Executive deficits were the strongest predictors of impairment on the global functioning indices, while memory and attention deficits were the strongest predictors of premorbid adjustment and social-vocational outcome 2 years after study entry. These results are generally consistent with those reviewed by Green (1986). Recent attention has focused on possible sex differences in neuropsychological functioning, and the implications of these differences for models of etiology and/or pathophysiology (Bilder, 1002; Goldstein et al., 1994; Lewine et al., 1990; 1996). After controlling for normal differences in motor speed (men are faster), we found no neuropsychological differences. We found, however, that within the patient group, women performed more poorly than men on the premorbid index. Because the premorbid index was loaded highly on certain tests (i.e., the information subtest of the WAIS-R) that show normal advantages for men over women (Kaufman, 1990), this finding may be an exaggeration of a normal sex difference, or fewer high-functioning women may have participated in testing. In either case, the results fail to support the hypothesis of more severe pathology among men with schizophrenia.

Handedness effects were prominent. Nondextral patients performed approximately 0.5 standard deviations below dextral patients, and the profile shape was similar. These results were not due to a subgroup with mixed-hand preference (Green et al., 1989). We previously noted in a study of WAIS-R results that only strongly dextral first-episode patients had IQ scores above the range of chronic patients (Green, 1996). These strongly dextral first-episode patients may be less likely to go on to a chronic course, or they may be more likely to show cognitive deterioration if and when their illness progresses. In either case, our results reveal a major association of nondextrality with global neuropsychological deficit. We previously hypothesized that this association might reflect a generalized failure in cerebral anatomic specialization (Bilder & Degreef, 1991), but so far there is little direct evidence linking neuropsychological deficits to absence or reversal of normal structural brain asymmetries.

To what extent can findings of neuropsychological deficit in schizophrenia be attributed to medication effects? This question is intractable outside the context of a controlled treatment trial (Bilder et al., 1992), but it remains important to assess relations of treatment with neuropsychological performance to guide future studies. We found significant correlations between antipsychotic dose and impairment on every neuropsychological scale, but higher dose also correlated with severity of hallucinations and the CGI score. We are no more inclined to suggest that antipsychotic drugs cause neuropsychological deficit than to suggest that antipsychotic drugs cause hallucinations.

The results indicate that patients with more severe, treatment-refractory symptoms had greater neuropsychological impairment and received higher doses. Further, given that the association of higher dose with neuropsychological impairment was reduced but not eliminated by statistically controlling for symptom severity, neuropsychological deficits may either prompt dose escalation or result from it. A similar pattern of results was observed with antiparkinsonian agents. In line with the literature (Spohn et al., 1989; Strauss et al., 1990), benztropine-equivalent dose correlated significantly with poorer performance on the memory scale, although it also correlated with poorer performance on executive, motor, and global scales, suggesting little specific effect of anticholinergic load. But benztropine-equivalent dose also correlated with global clinical ratings and items reflecting decreased spontaneous movement and poverty of speech. These findings suggest that patients with persistent akinesia and more neuropsychological deficit tend to receive more antiparkinsonian medication. Because memory deficit was still associated with benztropine-equivalent dose after controlling for extrapyramidal symptoms, there may be a separate anticholinergic effect on memory that deserves further study. Such studies must now consider the overall anticholinergic burden of regimens involving both antipsychotic agents and adjunctive treatments with high levels of anticholinergic activity.

References

Bentler, P.M., & Bonnett, D.G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bullettin*, 88, 588-606.

Bilder, R.M. (1985). Subtyping in chronic schizophrenia: clinical, neuropsychological, and structural indices of deterioration. *Dissertation Abstracts International*, 45 (11B), 3651.

Bilder, R.M. (1992). Structure-function relations in schizophrenia: brain morphology and neuropsychology. In: Walker, E.F., Dworkin, R.H., Cornblatt, B.A. (eds.), *Progress in Experimental Personality and Psychopathology Research*, vol 15. New York: Springer-Verlag, pp. 183-251.

Bilder, R.M., & Bates, J.A. (1994). Neuropsychological prediction of treatment response and outcome in schizophrenia. In: Gaebel, W., Awad, A.G. (eds.), *Prediction of Neuroleptic Treatment Outcome in Schizophrenia: Concepts and Methods*. New York: Springer-Verlag, pp. 99-110.

Bilder, R.M., Bogerts, B., Ashtari, M., Wu, H., Alvir, J.M., Jody, D., Reiter, G., Bell, L., & Lieberman, J.A. (1995). Anterior hippocampal volume reductions predict frontal lobe dysfunction in first episode schizophrenia. *Schizophrenia Research*, 17, 47-58.

Bilder, R.M., Degreef, G., Pandurangi, A.K., Rieder, R.O., Sackeim, H.A., & Mukherjee, S. (1988). Neuropsychological deterioration and CT scan findings in chronic schizophrenia. *Schizophrenia Research*, 1, 37-45.

Bilder, R.M., & Degreef, G. (1991). Morphologic markers of neurodevelopmental paths to schizophrenia. In: Mednick, S.A., Cannon, T.D., Barr, C.E., LaFosse, J.M. (eds.), *Developmental Neuropathology of Schizophrenia*. New York: Plenum, pp. 167-190.

Bilder, R.M., & Goldberg, E. (1987). Motor perseverations in schizophrenia. *Archives of Clinical Neuropsychology*, 2, 195-214.

Bilder, R.M., Lipschutz-Broch, L., Reiter, G., Geisler, S., Mayerhoff, D., & Lieberman, J.A. (1991). Neuropsychological deficits in the early course of first episode schizophrenia. *Schizophrenia Research*, 5, 198-199.

Bilder, R.M., Lipschutz-Broch, L., Reiter, G., Geisler, S.H., Mayerhoff, D.I., & Lieberman, J.A. (1992). Intellectual deficits in first-episode schizophrenia: evidence for progressive deterioration. *Schizophrenia Bulletin*, 18, 437-448.

Bilder, R.M., Mukherjee, S., Rieder, R.O., & Pandurangi, A.K. (1985). Symptomatic and neuropsychological components of defect states. *Schizophrenia Bulletin*, 11, 409-419.

Bilder, R.M., & Szeszko, P.R. (1996). Structural neuroimaging and neuropsychological impairments. In: Pantelis, C., Nelson, H.E., Barnes, T.R.E. (eds.), *The Neuropsychology of Schizophrenia*. New York: John Wiley & Sons, pp. 279-298.

Bilder, R.M., Turkel, E., Lipschutz-Broch, L., & Lieberman, J.A. (1992). Antipsychotic medication effects on neuropsychological functions. *Psychopharmacological Bulletin*, 28, 353-366.

Blanchard, J.J., & Neale, J.M. (1994). The neuropsychological signature of schizophrenia: generalized or differential deficit? *American Journal of Psychiatry*, 151, 40-48.

Cannon-Spoor, H.E., Potkin, S.G., Wyatt, R.J. (1982). Measurement of premorbid adjustment in chronic schizophrenia. *Schizophrenia Bulletin*, 8, 470-484.

Censits, D.M., Ragland, D.J., Gur, R.C., & Gur, R.E. (1997). Neuropsychological evidence supporting a neurodevelopmental model of schizophrenia: A longitudinal study. *Schizophrenia Research*, 24, 289-298.

Chapman, L.J., & Chapman, J.P. (1989). Strategies for resolving the heterogeneity of schizophrenics and their relatives using cognitive measures. *Journal of Abnormal Psychology*, 98, 57-366.

DeLisi, L.E., Tew, W., Xie, S., Hoff, A.L., Sakuma, M., Kushner, M., Lee, G., Shedlack, K., Smith, A.M., & Grimson, R. (1995). A prospective follow-up study of brain morphology and cognition in first-episode schizophrenic patients: preliminary findings. *Biological Psychiatry*, 38, 349-360.

Endicott, J., & Spitzer, R.L. (1978). A diagnostic interview: the schedule for affective disorders and schizophrenia. *Archives of General Psychiatry*, 35, 837-844.

Fish, B., Marcus, J., Hans, S.L., Auerbach, J.G., & Perdue, S. (1992). Infants at risk for schizophrenia: sequelae of a genetic neurointegrative defect: A review and replication analysis of pandysmaturation in the Jerusalem Infant Development Study. *Archives of General Psychiatry*, 49, 221-235.

Flor-Henry, P. (1983). *Cerebral Basis of Psychopathology*. Boston: John Wright-PSG.

Gold, S., Arndt, S., Nopoulos, P., O'Leary, D.S., & Andreasen, N.C. (1999). Longitudinal study of cognitive function in first-episode and recent-onset schizophrenia. *American Journal of Psychiatry*, 156, 1342-1348.

Goldberg, T.E., Weinberger, D.R., Berman, K.F., Pliskin, N.H., & Podd, M.H. (1987). Further evidence for dementia of the prefrontal type in schizophrenia? A controlled study of teaching the Wisconsin Card Sorting Test. *Archives of General Psychiatry*, 44, 1008-1014.

Goldman, R.S., Bates, J.A., Bilder, R.M., Reiter, G., Conley, J., Pappadopoulos, E., Obuchowski, M., Robinson, D., Alvir, J.M.A., Lieberman, J., & Schooler, N. (1999). Severity and selectivity of memory dysfunction in first-episode schizophrenia. *Schizophrenia Research*, 36, 130.

Goldstein, G. (1978). Cognitive and perceptual differences between schizophrenics and organics. *Schizophrenia Bulletin*, 4, 160-185.

Goldstein, J.M., Seidman, L.J., Santangelo, S., Knapp, P.H., & Tsuang, M.T. (1994). Are schizophrenic men at higher risk for developmental deficits than schizophrenic women? Implications for adult neuropsychological functions. *Journal of Psychiatric Research*, 28, 483-498.

Green, M.F., Satz, P., Smith, C., & Nelson, L. (1989). Is there atypical handedness in schizophrenia? *Journal of Abnormal Psychology*, 98, 57-61.

Green, M.F. (1996). What are the functional consequences of neurocognitive deficits in schizophrenia? *American Journal of Psychiatry*, 153, 321-330.

Gruzelier, J., Seymour, K., Wilson, L., Jolley, A., & Hirsch, S. (1988). Impairments on neuropsychologic tests of temporohippocampal and frontohippocampal functions and word fluency in remitting schizophrenia and affective disorders. *Archives of General Psychiatry*, 45, 623-629.

Gruzelier, J.H. (1991). Hemispheric imbalance: syndromes of schizophrenia, premorbid personality, and neurodevelopmental influences. In: Steinhauer, S.R., Gruzelier, J.H., Zubin, J. (eds.), *Handbook of Schizophrenia*, vol 5: *Neuropsychology, Psychophysiology and Information Processing*. New York: Elsevier, pp. 599-650.

Heaton, R., Paulsen, J.S., McAdams, L.A., Kuck, J., Zisook, S., Braff, D., Harris, J., & Jeste, D.V. (1994). Neuropsychological deficits in schizophrenics: relationship to age, chronicity, and dementia. *Archives of General Psychiatry*, 51, 469-476.

Heaton, R.K., Baade, L.E., & Johnson, K.L. (1978). Neuropsychological test results associated with psychiatric disorders in adults. *Psychological Bulletin*, 85, 141-162.

Hoff, A.L., Riordan, H., O'Donnell, D.W., Morris, L., & DeLisi, L.E. (1992). Neuropsychological functioning of first-episode schizophreniform patients. *American Journal of Psychiatry*, 149, 898-903.

Hoff, A.L., Sakuma, M., Wieneke, M., Horon, R., Kushner, M., & DeLisi, L.E. (1999). Longitudinal neuropsychological follow-up study of patients with first-episode schizophrenia. *American Journal of Psychiatry*, 156, 1336-1341.

Hoff, A.L., Wieneke, M., Faustman, W.O., Horon, R., Sakuma, M., Blankfeld, H., Espinoza, S., & DeLisi, L.E. (1998). Sex differences in neuropsychological functioning of first-episode and chronically ill schizophrenic patients. *American Journal of Psychiatry*, 155, 1437-1439.

Hollingshead, A.B., & Redlich, F.C. (1958). *Social Class and Mental Illness: A Community Study*. New York: John Wiley & Sons.

Kaufman, A.S. (1990). *Assessing Adolescent and Adult Intelligence*. Boston: Allyn & Bacon.

Kirkpatrick, B., Buchanan, R.W., McKenney, P.D., Alphs, L.D., & Carpenter, W.T. Jr. (1989). The Schedule for the Deficit Syndrome: An instrument for research in schizophrenia. *Psychiatry Research*, 30, 119-123.

Lewine, R.R., Gulley, L.R., Risch, S.C., Jewart, R., & Houpt, J.L. (1990). Sexual dimorphism, brain morphology, and schizophrenia. *Schizophrenia Bulletin*, 16, 195-203.

Lewine, R.R., Walker, E.F., Shurett, R., Caudle, J., & Haden, C. (1996). Sex differences in neuropsychological functioning among schizophrenic patients. *American Journal of Psychiatry*, 153, 1178-1184.

Lieberman, J.A., Alvir, J.M., Woerner, M., Degreef, G., Bilder, R.M., Ashtari, M., Bogerts, B., Mayerhoff, D.I., Geisler, S.H., Loebel, A., Levy, D.L., Hinrichsen, G., Szymanski, S., Chakos, M., Koreen, A., Borenstein, M., & Kane, J.M. (1992). Prospective study of psychobiology in first-episode schizophrenia at Hillside Hospital. *Schizophrenia Bulletin*, 18, 351-371.

Lieberman, J.A., Jody, D., Alvir, J.M., Ashtari, M., Levy, D.L., Bogerts, B., Degreef, G., Mayerhoff, D.I., & Cooper, T. (1993). Brain morphology, dopamine, and eye-tracking abnormalities in first-episode schizophrenia: prevalence and clinical correlates. *Archives of General Psychiatry*, 50, 357-368.

Lieberman, J.A., Jody, D.N., Geisler, S.H., Alvir, J.M., Loebel, A., Szymanski, S., Woerner, M., & Borenstein, M. (1993). Time course and biologic correlates of treatment response in first-episode schizophrenia. *Archives of General Psychiatry*, 50, 369-376.

Marcus, J., Hans, S.L., Mednick, S.A., Schulsinger, F., & Michelsen, N. (1985). Neurological dysfunctioning in offspring of schizophrenics in Israel and Denmark: A replication analysis. *Archives of General Psychiatry*, 42, 753-761.

Mayerhoff, D.I., Loebel, A.D., Alvir, J.M., Szymanski, S.R., Geisler, S.H., Borenstein, M., & Lieberman, J.A. (1994). The deficit state in first-episode schizophrenia. *American Journal of Psychiatry*, 151, 1417-1422.

Medalia, A., Gold, J.M., & Merriam, A.E. (1988). The effects of neuroleptics on neuropsychological test results of schizophrenics. *Archives of Clinical Neuropsychology*, 3, 249-271.

Mednick, S.A., Mura, E., Schulsinger, F., & Mednick, B. (1974). Perinatal conditions and infant development in children with schizophrenic parents. In: Mednick, S.A., Schulsinger, F., Higgins, J., Bell, B. (eds.), *Genetics, Environment, and Psychopathology*. Amsterdam: Elsevier, pp. 231-248.

Meng, X., Rosenthal, R., Rubin, D.B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, 111, 172-175.

Mohamed, S., Paulsen, J.S., O'Leary, D., Arndt, S., & Andreasen, N. (1999). Generalized cognitive deficits in schizophrenia: A study of first-episode patients. *Archives of General Psychiatry*, 56, 749-754.

Neumann, C.S., Grimes, K., Walker, E.F., & Baum, K. (1995). Developmental pathways to schizophrenia: behavioral subtypes. *Journal of Abnormal Psychology*, 104, 558-566.

Nuechterlein, K.H., Gitlin, M.J., & Subotnok, K.L. (1995). The early course of schizophrenia and long-term maintenance neuroleptic therapy. *Archives of General Psychiatry*, 52, 203-205.

Oldfield, R.C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9, 97-113.

Robinson, D., Woerner, M.G., Alvir, J.M., Bilder, R., Goldman, R., Geisler, S., Koreen, A., Sheitman, B., Chakos, M., Mayerhoff, D., & Lieberman, J.A. (1999). Predictors of relapse following response from a first episode of schizophrenia or schizoaffective disorder. *Archives of General Psychiatry*, 56, 241-247.

Robinson, D.G., Woerner, M.G., Alvir, J.M., Geisler, S., Koreen, A., Sheitman, B., Chakos, M., Mayerhoff, D., Bilder, R., Goldman, R., & Lieberman, J.A. (1999). Predictors of treatment response from a first episode of schizophrenia or schizoaffective disorder. *American Journal of Psychiatry*, 156, 544-549.

Saykin, A.J., Gur, R.C., Gur, R.E., Mozley, P.D., Mozley, L.H., Resnick, S.M., Kester, D.B., & Stafiniak, P. (1991). Neuropsychological function in schizophrenia: Selective impairment in memory and learning. *Archives of General Psychiatry*, 48, 618-624.

Saykin, A.J., Shtasel, D.L., Gur, R.E., Kester, D.B., Mozley, L.H., Stafiniak, P., & Gur, R.C. (1994). Neuropsychological deficits in neuroleptic naive patients with first-episode schizophrenia. *Archives of General Psychiatry*, 51, 124-131.

Schachter, S.C., Ransil, B.J., & Geschwind, N. (1987). Associations of handedness with hair colour and learning disabilities. *Neuropsychologia*, 25, 269-276.

Spitzer, R.L., Endicott, J., & Robins, E. (1977). *Research Diagnostic Criteria (RDC) for a Selected Group of Functional Disorders*, 3rd. ed. New York: New York State Psychiatric Institute, Biometrics Research.

Spitzer, R.L., Endicott, J., & Robins, E. (1978). Research Diagnostic Criteria rationale and reliability. *Archives of General Psychiatry*, 35, 773-782.

Spohn, H.E., & Strauss, M.E. (1989). Relation of neuroleptic and anticholinergic medication to cognitive functions in schizophrenia. *Journal of Abnormal Psychology*, 98, 367-380.

Squire, L.R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, 253, 1380-1386.

Strauss, M.E., Reynolds, K.S., Jayaram, G., & Tune, L.E. (1990). Effects of anticholinergic medication on memory in schizophrenia. *Schizophrenia Research*, 3, 127-129.

Sullivan, E.V., Shear, P.K., Zipursky, R.B., Sagar, H.J., & Pfefferbaum, A. (1994). A deficit profile of executive, memory, and motor functions in schizophrenia. *Biological Psychiatry*, 36, 641-653.

Szeszko, P.R., Bilder, R.M., Lencz, T., Ashtari, M., Goldman, R.S., Reiter, G., Wu, H., & Lieberman, J.A. (1999). Reduced anterior cingulate gyrus volume correlates with executive dysfunction in men with first-episode schizophrenia. *Schizophrenia Research*, 45, 680-686.

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.

Walker, E., & Levine, R.J. (1990). Prediction of adult-onset schizophrenia from childhood home movies of the patients. *American Journal of Psychiatry*, 147, 1052-1056.

Willson, D.F. (1997). *The Nature of Accelerated Forgetting Rates in Schizophrenia* (doctoral dissertation). Jamaica, NY, St. John's University, Department of Psychology.

The cognitive science of metaphor from philosophy to neuropsychology

Tim Rohrer

Institute for Neural Computation, University of California, San Diego

rohrer.pf@darkwing.uoregon.edu

Abstract

In this paper I review some of the theoretical issues surrounding metaphor, and trace them through the context of the cognitive neuroscience debate. Metaphor, like all figurative language, has been usually explained as a secondary linguistic process which takes place as a function taking place on literal language. However this explanation does not fit well with some of the recent work on right hemisphere processing of language or recent cognitive studies, both of which suggest that the figurative and literal language are processed simultaneously and share much structure. In seeking ways to operationalize the Lakoff and Johnson view of metaphor as a constitutive cognitive phenomenon, I begin to spell out what kinds of theoretical predictions the Lakoff-Johnson model would make on the neurophysiological levels cognitive investigation. I conclude by offering some rudimentary thoughts on possible proposals for further investigation using these methods.

Keywords: Metaphor; Cognitive Processes; Right Brain

1. Introduction and theoretical overview

> *Whatever the skill employed in thought-that of logic, mathematics, language, spatial or musical symbols-we must not forget that it is driven by the Jamesian processes, undergoes flights and perchings, is susceptible to great variations in attention, and in general, is fueled by metaphorical and metonymic processes.* (Gerald Edelman, *Bright Air, Brilliant Fire*, p. 174)

Though it has long been recognized that most of our everyday uses of language involve metaphor, with a few notable exceptions (such as Winner &

Garner, 1977) the enterprise of cognitive science has largely ignored the investigation of figurative language in favor of investigating literal language until recently. Of the many factors which contributed to the paucity of research on figurative language comprehension, the *instantiation hypothesis* is perhaps the most onerous. The instantiation hypothesis argued that reason, intelligence and minds were *substrate neutral*, that is, independent of any specific embodiment, so long as the body was a algorithmic device. The computational device of choice was a serial processor driven digital computer. Since the mathematician Alan Turing proved that all digital computers were in principle reducible to recursive elaboration of a finite state algorithm (a "Turing machine"), minds (and mental "processes") were in principle reducible to finite state algorithms. Finite state algorithms have a peculiar literal quality in that their variables are either true or false with no admixture of truth or falsity permitted. Upon this view the fundamental problem of language comprehension was determining how the brain's representations and the world literally matched up: the world outside the brain was thought to be represented symbolically inside the brain by a series of finite states. If the mind was the kind of software program running on the that kind of hardware, the lack of attention given to figurative language comprehension results from an obvious source: it would be a mere afterthought to solving the problem of literal language comprehension, since Turing had proven that all computational processes must ultimately decompose into finite states-that is, into literality. The mantra of this dogma was clear: solve how language represents and the problems posed by figurative language will inevitably solve themselves.

Unfortunately for proponents of the instantiation hypothesis, many of the recent findings in cognitive science are motivated by its antithesis: the *embodiment hypothesis*. The embodiment hypothesis argues that minds are fundamentally not disembodied algorithmic processes like a computer program, but are instead constituted and constrained by the kinds of organization reflected in the biological, anatomical, biochemical, and neurophysiological characteristics of the body and the brain. While both hypotheses share the materialist assumption that higher-level processes, such as abstract thought, language comprehension and the like, are built up out of lower-level processes, the embodiment hypothesis explicitly denies the substrate neutrality claim of the instantiation hypothesis proponents. Rather than processes which manipulate symbols in a finite-state fashion, the study of mind as a biological and neurophysiological enterprise requires thinking about the mind in a series of interwoven levels of investigation, many of which appear to behave analogically and frequently exhibit a kind of adaptive plasticity not found in digital computers. A good cognitive theory requires a kind of vertical convergence across a number of different levels of investigation ranging from philosophy and cognitive psychology through neuropsychology to neuroanatomy and neurochemistry.

Figurative language comprehension has robust connections with the embodiment hypothesis, especially in the area of metaphor. As George Lakoff (1987) and Mark Johnson (1987) have argued, our ordinary use of language is largely structured by metaphoric and metonymic principles which exhibit a directionality. Human beings systematically characterize abstract ideas-thoughts, religious beliefs, political and ethical situations-in terms of bodily movements and bodily functions, for example. The primary claim of their position is that these metaphors and the directionality are not arbitrary, but instead are a natural outgrowth of the manner in which our minds are constituted. In this paper I seek to tie together threads from the philosophical, cognitive, and neurophysiological levels. In operationalizing the Lakoff-Johnson hypothesis I have made use of Posner's schematization of cognitive science (Table 1) as a framework for seeking a vertically integrated cognitive theory.

Table 1. Overview of the proposed operationalization of the conceptual mapping hypothesis on metaphor in relation to Posner's conceptual framework for characterizing the levels of investigation in cognitive science

Conceptual Metaphor Theory	Level of Investigation	Methods of Study
Tasks		
Understanding a metaphor	Cognitive Systems	Verbal report, textual analysis, discourse analysis
Cuing and retrieval for a conceptual metaphor	Mental operations	Cognitive studies examining RT
Facilitation of related information, inference generalizations	Performance Domain	RT in priming experiments with polysemic words, ERP studies of the time course involved
Activation in somatosensory, auditory, and visual processing areas when processing metaphorically constituted abstract concepts	Neural Systems	Lesion, FMRI, PET, sometimes in combination with ERP
Intraneural connections from visual, auditory, somatosensory region to anguage areas	Cellular	Electrocellular stimulation, and anatomical methods, supplemented by computer simulations

2. Literature Review: philosophical work and cognitive systems

Most philosophers have generally argued that figurative language involves tricks or plays on the literal. John Searle (1979), for example, argued that metaphor is a simply a roundabout way to express literal semantics. According to Searle's model, all utterances would be processed as literal utterances first. Only once the mind was unable to find a literal meaning for an expression would the utterance be sent to a special non-literal processing center for decoding. Searle's view necessarily entails that the comprehension of a metaphor assumes the metaphorical expression will eventually be decomposed into a literal paraphrase.

On the face of it Searle's model has much going for it, for it fits with our common sense experience of understanding language and speech. However that common sense understanding fundamentally confounds the distinction between literal and figurative speech with another distinction: that between attended and automatic processes. In our common sense understanding we think of metaphors as unusual utterances which require our attention, such as we encounter in poetry or theater. We think of a metaphor as something jarring which interrupts our ordinary way of thinking about the world and challenges us to enter into another way of thinking. However, Lakoff and Johnson have shown that much of our everyday language, including what we would ordinarily call literal language, is structured by conventional metaphors. In other words, the bulk of our metaphoric processing is automatic-only some metaphors require attentional processing. Couple their observations with the fact that some forms of language which we would ordinarily consider literal (such as scientific reports) also requires attended processing even though most forms of literal language (such as social interactions, reading a newspaper and the like) are also processed fairly automatically, and Searle's model of how we understand metaphor begins to pale. This realization suggests two different kinds of theoretical models: one in which both literal and metaphorical processing are done in parallel and another in which both literal and metaphorical processes are largely the same process.

A number of cognitive methods can be applied to help adjudicate between these three theoretical models. As Searle's model is a sequential hypothesis, it requires that a metaphorical utterance first be processed as if it were literal, judged non-literal and then alternate strategies are employed to decode the utterance into literal. Presumably all these actions take place in real time, which would imply that metaphorical utterances take longer to understand than literal utterances. However, many studies of figurative language-on idioms, proverbs and metaphor-show this is not the case (see Gibbs

1994), provided there is sufficient context. Ortony, Schallert, Reynolds and Antos (1978) measured the time it took subjects to comprehend literal v. metaphorical sentences at the end of long and short contexts, and reported no difference for long contexts but that metaphorical sentences took significantly longer to read in short contexts. Janus and Bever (1985) tracked eye movements and compared the amount of time people focused on the target sentences. Subjects again responded as quickly in the long context condition. These results contraindicate the Searlean sequential model of metaphor processing; additionally, the possibility that metaphorical contexts are 'chunked' and processed as semantic units is contraindicated by the differing results in the long v. short context conditions. (Presumably if metaphors were chunked in a fashion similar to lexemes they would have been retrieved in nearly equal times for either the short or long context conditions – this was clearly not the case). Finally, it should be noted as a caveat that other studies (Blasko & Connine, 1993) show that highly familiar metaphors are understood more quickly than novel metaphors-in other words, these processing time studies must be taken cautiously as they show practice effects.

While these type of cognitive studies adequately contraindicate the Serlean model of metaphorical processing, they have not yet proved as revealing in determining whether literal and figurative language processing are parallel processes or are largely the same process. With respect to idiomatic language Gibbs (1980; 1986) has shown that subjects take less time to read idiomatic phrases when the context supports an idiomatic interpretation than the same phrases in contexts supporting a literal interpretation. The differing results with respect to context suggest that the literal and figurative language comprehension processes operate in parallel and the context primes one process or the other. Similar results have been obtained with regard to metaphor (summarized in Gibbs, 1994). Yet subjects seem unable to ignore a figurative meaning even when instructed to focus exclusively on the literal context (Keysar, 1989). Keysar investigated whether metaphor and literal interpretations of a context would produce a Stroop-like interference effect when a target was, for example, both metaphorically true and literally false. Reaction times increased in the invalid condition but decreased in the valid condition, suggesting that these processes may well at least share component subsystems. Ultimately cognitive methods alone may not be able to decisively resolve the parallel and same process debate because that question may be better posed at other levels of investigation than the cognitive systems level.

Finally, there has recently been an interesting exchange between the Lakoff-Johnson-Gibbs approach and the largely complementary Glucksberg-Keysar-McGlone approach. Along with Lakoff and Johnson, Gibbs (1992) claims that long-term memory is structurally organized by prototypes extended by metaphoric and metonymic principles called conceptual metaphors or conceptual mappings. A conceptual metaphor, such as LOVE IS A JOURNEY, is constituted by a mapping between areas of the brain, such as between affect and sensorimotor areas. In the appropriate context, most conventional metaphoric expressions, such as *we're at a crossroads* or *our marriage is on the rocks*, access these conceptual metaphors from long-term memory. Glucksberg, Keysar and McGlone (1990) argue for a *class inclusion* view in which some metaphoric expressions build up an ad-hoc category in working memory rather than accessing conceptual metaphors from long-term memory, even when it would be expected that they would draw on a conceptual metaphor. Gibbs (1992; 1994) argues that the class-inclusion view requires an understanding of metaphor in which each metaphorical expression creates a unique or novel mapping in working memory. In their reply Glucksberg, Keysar and McGlone (1992) suggest that only some cases require the development of ad-hoc category, citing a brief initial experiment in which subjects were given metaphorical with minimal context and asked to paraphrase them.

3. Literature Review: neurophysiological investigations of metaphor

Some of the strongest evidence against a purely parallel processing model for figurative and literal language comprehension comes from a number of studies on right hemisphere-damaged (RHD) patient populations. One possible localization of these parallel processes would suggest that they could be lateralized with respect to brain hemispheres. In an early study often cited as establishing that metaphoric processing is right hemisphere dependent, Winner and Gardner (1977) compared left hemisphere-damaged (LHD) aphasic patients, RHD patients, bilaterally damaged patients and a non-neurological group. Participants were presented with a figurative sentence such as "he has a heavy heart" and asked to perform two tasks. One task involved matching the sentence to one of four pictures, one of which literally represented the sentence (i.e., a man carrying a large heart), an appropriate metaphoric interpretation (a crying man), a salient quality depicted by the metaphoric adjective (a 500 lb weight) and one which illustrated the noun (a large heart). In the second task patients were asked to verbally explain their choices. While RHD patients selected the metaphoric picture much less frequently than LHD patients or the control group in the first task, in the second task RHD patients were able to verbalize their choices using figurative language whereas LHD patients were unable to explain their 'correct' choices verbally. Winner and Gardner argue that these results show that the effective interaction of the hemispheres was important in appreciating figurative meaning. Though this study clearly shows that the figurative and literal language comprehension processes are not parallel processes in differing hemispheres, the study also indicates that at least some subcomponents of language comprehension concerned with metaphoric processing are RH dependent. In short, the RH makes a positive contribution to figurative language processing. This kind of lesion evidence further suggests that parallel processing of literal and figurative language comprehension is too crude a level of analysis to be a useful description at the level of localization, suggesting that the figurative and literal language processing share at least some but not all components.

This conclusion was underscored in a related study by Brownell et al. (1990) which also showed that RHD patients were more impaired in pairing a word with a metaphoric synonym than LHD patients. In addition to providing further support for the idea that the right hemisphere makes a unique contribution to figurative language processing, they also report that LHD patients were affected by the degree of semantic similarity between the target and its synonym in the metaphoric-but not the literal-condition. This finding is intriguing because it suggests that the right and left hemispheres may code semantic information in different ways. Work on RH contributions to semantic priming (summarized in Beeman et al., 1993) generally suggests that semantic memory is either stored differently in the RH or processed differently in the RH. Beeman argues for the latter, suggesting that the RH and LH do operate in parallel with respect to semantic processing, with the RH processing the information in a coarser fashion and the LH in a fine fashion. The coarse-fine distinction is drawn from neural network modelling of vision which suggests that networks which use larger receptive fields code information more coarsely, thereby maximizing responsivity to precisely localizing a source of continuous input. According to this hypothesis the coarser semantic fields of the RH would operate in parallel with the more finely grained semantic fields of the left hemisphere. Figurative language comprehension would then entail a RH activation strong enough to influence LH hemisphere processing.

If Beeman's theory can be adapted to understanding the phenomenon of conceptual metaphor, the context effect observed in the cognitive studies would presumably be instantiated as a build-up of strong patterns of activation in the RH. This explanation is also consonant with the Stroop-like interference effect that Keysar (1989) noted in which the availability of metaphoric interpretations slowed the RT of literal interpretations. By positing a shared semantic network which differs only in the scope and weighting of the neural nets, this theoretical model nicely captures both the shared componentry evidenced by the interference between literal and figurative processing and the parallel processing suggested by the decreased reaction times in the long context condition. Rather than architecturally separate modules for literal and figurative processing, this theoretical model posits a literal-metaphorical continuum instantiated by a varying the organization of neural networks. However, the conceptual metaphor model seems to argue against a strict localization of metaphor in th RH as it requires activation in related sensorimotor areas (i.e., visual metaphors should activate visual areas, etc.). This may prove to be a significant incompatibility with Beeman's hypothesis.

4. Proposals for further investigations

There remain, however, unanswered questions about the role of conceptual metaphors as largely automatic (unattended) processes in the brain. Lakoff (1994) argues that "the system of conventional conceptual metaphor is mostly unconscious, automatic and available to users with no noticeable effort". If conceptual metaphors are largely automatic and metaphoric processing is largely instantiated by coarse semantic fields in the RH, the RH should show a strong and growing activation pattern in the long context condition of the cognitive studies on metaphor. This hypothesis could be investigated using FMRI methodology by reproducing the cognitive studies on metaphor and rather than measuring the RT needed to comprehend a sentence, look for a growing pattern of activation in the RH. Imaging techniques could compare early, middle and late activations involved in the comprehension of the passage. I believe the appropriate subtractions would be the event images minus an image of the subject at rest (either reading nonesense text or 'literal' text). Presumably these subtractions would show greater RH activation late in the passage. If ERPs were used in conjunction with a neuroimaging method, a clearer picture of the time course might provide us with insight about the rate of the expected build-up in the RH.

Another line of inquiry to investigate would focus on the notion of a conceptual metaphor. The conceptual metaphor hypothesis proposes that there will be interconnections between semantic areas and areas tightly related to sensory input. A passage which uses visual metaphors, for example, should activate pathways connected to the visual system and to some degree the areas themselves. A FMRI study could be designed which checked for activation in the visual, auditory, somatosensory and motor cortices after reading or hearing a passage with strong visual, auditory, tactile or bodily movement metaphors. The relevant subtraction would change in each modality, but the basic approach would be to compare a loaded condition minus an unloaded condition; e.g., for the visual modality an image after hearing a passage with visual metaphors minus an image of passive listening to a visual-metaphor neutral text. One would expect to see activations. If there was no such activation, it would be substantial evidence contraindicating the conceptual metaphor view. However, all of these suggestions are subject to the caveat that the activation should be of large enough scale to be picked up by the resolution of the neuroimaging technique. On the neuroanatomical level, pathways should be observable between associated areas.

Finally it strikes me that the differences between the class inclusion view and the conceptual metaphor view may also lend themselves to investigations at the neurophysiological level, but I am not clear on how the distinction between long term and working memory can be made testable at this time, and that appears to be the key bone of contention for the experimental design.

5. Conclusion

I have provided an initial attempt at linking the conceptual metaphor hypothesis to part of the recent neurophysiological literature on metaphor. In tracing theoretical issues through this debate it has become clear that the differing levels of investigation have historically had differing concerns which are only beginning to be translatable into a vertically integrated cognitive theory which encompasses levels from biology and anatomy through neurophysiology and cognitive studies. I have focused on the relationship between neurophysiology, cognitive studies and philosophy in this paper. Finally, I've proposed some avenues of neurophysiological investigation which are theoretically significant given recent philosophical and cognitive work.

References

Beeman, M., Friedman, R.B., Grafman, J., Perez, E., Diamond, S., & Lindsay, M.B. (1994). Summation Priming and coarse semantic coding in the right hemisphere. *Journal of Cognitive Neuroscience*, 6, 26-45.

Blasko, D., & Connine, C. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology: Learning Memory, and Cognition*, 19, 295-308.

Brownell, H.H., Simpson, T.L., Bihrle, A.M., Potter, H.H., & Gardner, H. (1990). Appreciation of metaphoric alternative word meanins by left and right brain-damaged patients. *Neuropsychologia*, 28, 375-383.

Edelman, G.M. (1992). *Bright Air, Brilliant Fire: on the Matter of Mind*. New York: Basic books.

Gibbs, R. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory & Cognition*, 8, 449-456.

Gibbs, R. (1986). Skating on thin ice: literal meaning and understanding idioms in conversation. *Discourse Processes*, 9, 17-30.

Gibbs, R. (1992). Categorization and metaphor understanding. *Psychological Review*, 99, 572-577.

Gibbs, R. (1994). *The Poetics of Mind: Figurative Thought, Language and Understanding*. Cambridge, Eng.: Cambridge University Press.

Glucksberg, S., Keysar, B., & McGlone, M. (1990). Metaphor understanding and accessing conceptual schema: Reply to Gibbs (1992). *Psychological Review*, 99, 578-581.

Janus, R., & Bever, T. (1985). Processing metaphoric language: An investigation of the three stage model of metaphor comprehension. *Journal of Psycholinguistic Research*, 14, 473-487.

Johnson, M. (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination and Reason*. Chicago: University of Chicago Press.

Keysar, B. (1989). On the functional equivalence of literal and metaphorical interpretations in discourse. *Journal of Memory and Language*, 28, 375-385.

Lakoff, G. (1987). *Women, Fire and Dangerous Things*. Chicago: University of Chicago Press.

Lakoff G. (1994). *The contemporary theory of metaphor*. In: Ortony, A. (ed.), *Metaphor and Thought*, 2nd. ed. Cambridge: Cambridge University Press.

Ortony, A., Schallert, D., Reynolds, R., & Antos, S. (1978). Interpreting metaphors and idioms: Some effects of context on comprehension. *Journal of Verbal Learning and Verbal Behavior*, 17, 465-477.

Searle, J. (1979). Metaphor. In Ortony, A. (ed.), *Metaphor and Thought*. Cambridge: Cambridge University Press.

Winner, E., & Gardner, H. (1977). The Comprehension of metaphor in brain-damaged patients. *Brain*, 100, 717-729.

The iconic representation of metaphor

An event-related potentials (ERPs) analysis of figurative language

Michela Balconi - Serafino Tutino

Department of Psychology, Catholic University of Milan, Milan, Italy

michela.balconi@unicatt.it

Abstract

The aim of the study is to explore the iconic representation of frozen metaphor. Starting from the dichotomy between the pragmatic models, for which metaphor is a semantic anomaly, and the direct access models, where metaphor is seen as similar to literal language, the cognitive and linguistic processes involved in metaphor comprehension are analyzed using behavioural data (RTs) and neuropsychological indexes (ERPs). 36 subjects listened to 160 sentences equally shared in the variables content (metaphorical vs literal) and congruousness (anomalous vs not semantically anomalous). The ERPs analysis showed two negative deflections (N3-N4 complex), that indicated different cognitive processes involved in sentence comprehension. Repeated measures ANOVA, applied to peak amplitude and latency variables, suggested in fact N4 as index of semantic anomaly (incongruous stimuli), more localized in posterior (Pz) area, while N3 was sensitive to the content variable: metaphor sentences had an ampler deflection than literal ones and posteriorly distributed (Oz). Adding this results with behavioral data (no differences for metaphor vs literal), it seems that the difference between metaphorical and literal decoding isn't for the cognitive complexity of decoding (direct or indirect access), but for its representation format, which is more iconic for metaphor (as N3 suggests).

Keywords: Metaphor; Iconic Representation; Neuropsychology; EEG; Event-related Potentials

1. Introduction

The sentence "The fighters are lions" is a *common* metaphor, easy to understand by *common* people, that involves cognitive and linguistic processes.

The metaphor was defined as a process of correlation between two different domains: a structural and semantic transfer from a conceptual domain to another (Lakoff & Johnson, 1980). This definition implicates two levels of interest: the semiotic interpretation level (that concerns the question of meaning) and the cognitive level (the process involved in metaphor decoding).

For the first, if metaphor is the transfer of a property or concept (*vehicle*) to another, not conventional context (*topic*), then vehicle has an other meaning from the usual one, or rather an nonliteral meaning.

Secondly, metaphor doesn't seem to be only a "figure of speech", but even a *modus operandi* of our mind, where some specific processes are involved, as conceptualization, semantic memory activation, the use of iconic representation and semantic attribution, language construction and inference processes (Grady, 2005).

Thus, the linguistic and cognitive aspects are both important to explore metaphor comprehension, where the second aspect is a main issue of recent studies. Previous contributions have focused two dichotomic axes of analysis, even though in many cases such dichotomy has not been opportunely conceptualized:

1) A first question is the opposition between literal vs non literal meaning: is metaphor a "different" or "other" meaning in comparison to the literal, conventional one? (Glucksberg, 2003).

For the pragmatic approach metaphor is usually conceptualized as a shifting from "standard" understanding: some models investigated the semantic status of the metaphor, considering it as an example of linguistic anomaly (Grice, 1975). More specifically, the attribution of metaphoric meaning is possible only after the failure of the literal, standard interpretation and the consequent recognition of semantic anomaly (Bonnaud et al., 2002; Gibbs & Gerring, 1989). A second approach treats the question from a communicative point of view, with the distinction among *direct* and *indirect* elaboration: for some authors the metaphoric meaning decoding required an "indirect" process of analysis (Searle, 1979). In this perspective the inhibition of literal meaning was conceptualized as the key for the process of comprehension of metaphor (Glucksberg, 2003; McGlone & Manfredi, 2001). Bonnaud and coll. (2002) point out as central questions a preliminary lexical control process, a mechanism to elaborate sentences, for which metaphors gains a special status in the semantic memory.

Some behavioural indexes confirm the hypothesis of standard pragmatic model (as the increasing of the response time in reading metaphoric sentences in relation to literal ones), underlying that a literal decoding is a first, necessary step in sentence elaboration and so metaphor comprehension is realized only after it.

Nevertheless, not all results confirm the predominance of the literal on metaphoric meaning (Wolff & Gentner, 2000; Gibbs, 1994). A second set of research, in fact, lean towards the immediacy in decoding metaphor: the metaphorical meaning is automatic and direct (Kazmerski, Blasko & Dessalegn, 2003; Pynte et al., 1996). Some behavioural studies found evidences agree to this theory, showing similar RT (response time) for metaphoric and literal decoding, where the figurative meaning is considered immediate and direct, the main meaning to be elaborated, able to interfere with literal interpretation (Glucksberg, 2003; Iakimova et al., 2005; McElree & Nordlie, 1999; Ortony, 1979; Pynte et al., 1996; Tartter et al., 2002).

2) The second issue concerns the usual definition of metaphor as figurative language, for its specific ability to activate a pictorial representation format.

In fact, some models postulate that exist two different representational format for the elaboration of the word: an iconic elaboration, typical of metaphoric meaning, and a linguistic elaboration (Paivio, 1991; West & Holcomb, 2002): the dichotomy here is set along the literal vs iconic representation. Kazmerski and coll. (2003) correlated the ability in decoding metaphor and the subjects' skill to produce images: they found that the activation of a pictorial code or of a subsystem of the semantic memory is responsible of different subjects' performances. This subsystem would process iconic information in a qualitatively different way from the subsystem involved for linguistic information processing (Gentner & Wolff, 1997; Kintsch, 1988; Hamm et al., 2002; Paivio, 1991).

The second issue regards therefore the representational format of the figurative meaning (Kazmerski, Blasko & Dessalegn, 2003).

1.1. The neuropsychological approach: ERPs (event-related potentials) research

The contribute of neuropsychology was important to explore the neuropsychological structures, the linguistic and cognitive mechanisms involved in metaphor comprehension; in particular the electroencephalographic indexes of the event-related potential (ERPs) gave important information to investigate the relation between literal and figurative meaning decoding.

In fact, giving a good level of control of the cognitive phenomena, with a high temporal resolution, the ERPs allow to monitor the cognitive process sequentially, being the EEG modifications related to the qualitative changes of the processes implicated (Handy & Khoe, 2005; Rugg & Coles, 1995).

N400 is a recent ERP marker of sentence elaboration, a peak around 400 ms after the onset of the stimulus that is correlated with the presence of a detected anomaly, or a violation of a an attended information, in meaning comprehension. In general, N400 was an index that regards the cognitive level of analysis: when the cognitive system needs for the integration and the updating of the meaning of a stimulus within an anomalous or incongruent semantic context (Balconi & Lucchiari, 2005; Balconi & Pozzoli, 2003; 2004; Kutas & Hillyard, 1980; Rugg & Coles, 1995). It was widely used in metaphor research, showing contrasting results, even as consequence of different interpretative approaches adopted. From one side, Coulson and Van Petten (2002) found a N400 of greater intensity for the metaphoric sentences in comparison with the literal ones, but a meaningful differentiation was found even between metaphorical sentences and literal sentences that were semantically false. The authors suggested a broad similarity between metaphorical and literal processing: they would implicate the same mechanisms and the same timing elaboration. The increased peak amplitude for N400 was explained as the request of greater cognitive complexity for metaphoric meaning. The metaphoric and literal meaning would be therefore similar for the processes involved, but they would have only a difference for the cognitive gradient of difficulty, with a progressive increase in understanding metaphors.

Pynte et al. (1996) found a variation of gradual intensity of N4 in three conditions: at first literal, then conventional metaphoric and finally non conventional metaphoric sentences. They argued that other elements play a fundamental role in sentences comprehension: context and familiarity are relevant in the figurative language decoding, able to make the metaphoric meaning similar to literal meaning in the process of comprehension.

Kazmerski et al. (2003) compared metaphor, literal and scrambled sentences with the aim to explore the question of semantic anomaly. They found that is more difficult to evaluate as literally false a metaphor in comparison to a scrambled sentence: this would suggest that figurative language doesn't ask for a higher semantic integration, as it happens for scrambled utterances.

Moreover, Tartter and coll. (2002) didn't find differences for the N400 effect among literal and metaphoric meaning, but rather they showed differences in the early phase of elaboration (N200 effect), that could indicate the existence of different syntactic processes among the two conditions.

Finally, Iakimova et al. (2005) submitted literal and metaphorical sentences in a clinical domain, to normal and pathological subjects (schizophrenic), founding more difficult for all sentences in the comprehension of schizophrenic subjects, but not in relation to the metaphoric vs literal utterances. The authors suggested that these results are in line with the model of the "direct access" of the metaphoric meaning.

Besides, it is possible to compare the two conditions of decoding considering a second ERP variation, the N300 effect, that was found in case of an iconic representation of the meaning (Eddy et al., 2006; Federmeier & Kutas, 2001; Hamm et al., 2002; Laeng et al., 2003; Large et al., 2004). Holcomb et al. (1999) found this negative deflection for pictorial stimuli or linguistic stimuli with a high iconic value, as for images or lexical stimuli with an increased degree of pictorial feature. West and Holcomb (2002) stated that this variation was correlated to the process of transposition of the iconic format in a linguistic representation.

In synthesis, both the ERPs components, one related to the semantic elaboration of the stimulus (N400), the other related to the pictorial representational format (N300), can inform about the nature of the metaphoric decoding.

2. Objectives and hypotheses

The aim of this research is the comparison of the elaboration of the two conditions of linguistic and metaphoric decoding, considering both the semantic elaboration level and the representational format of the meaning. We will analyzed two distinct ERP deflections (N300 and N400 peaks) and behavioural ones (variations in the response time, RT) to the sentences. Specifically:

1. We will compare the literal and metaphor meaning comprehension to analyze the semantic attribution process of the meaning, paying attention to a potential detection of semantic anomaly in metaphorical meaning comprehension. In case of a direct access of the metaphoric meaning, in fact, the N4 deflection should have a similar profile for the two conditions. On the contrary, we expect to find a clear distinction between not anomalous (congruous) vs anomalous (incongruous) sentences in N4 amplitude. In other terms, the presence of semantically anomalous stimuli (incongruous sentences) should be marked by qualitatively different processes from those activated by semantically congruous (as metaphoric as literal) sentences (congruous sentences).
2. Also, peak latency can inform about the nature of the cognitive processes involved in the literal/metaphoric conditions and congruence/incongru-

ence dichotomy. Meaningful differences in the temporal onset of the N400 effect could suggest quantitatively different processes in sentence comprehension, indicating variances in cognitive complexity. A greater latency could be index of an increase in the cognitive complexity, mirror of longer times of elaboration.

3. Behavioral data give further information about the question of greater complexity in metaphor elaboration: an increase for figurative meaning decoding could suggest the necessity of semantic integration in comparison to literal elaboration. Contrarily, a substantial homogeneity of RTs would indicate similar degrees and levels of elaboration, confirming the model of the direct access of metaphorical sentences.
4. Moreover, figurative language could show a typical and different representational format of the meaning: an iconic format marked by specific ERP indexes. In line with this hypothesis, the presence of the negative component N3 will be observed, that was previously found in relation to stimuli with an iconic representational format.
5. Finally we will consider a possible ERP localization effect and in particular for literal vs metaphorical meaning. Localization effect could indicate a cortical specialization for the elaboration of the metaphorical vs. literal meaning, with a different contribution of the one or the another hemisphere. Some studies noticed that the right hemisphere, in general, is more involved then the left for the pragmatic components of the meaning (Beeman & Chiarello, 1998; Newman et al., 2003) as for metaphoric, ironic or sarcastic sentences (Bottini et al., 1994; Giora, 2003; Papagno, Oliveri & Romero, 2002). On the contrary, some studies didn't find these ampler activation of right hemisphere (Stringaris et al., 2006; Kacinik & Chiarello, 2007), leaving open the debate about the contribution of the two hemispheres in figurative language elaboration (Papagno & Carporali, 2007).

3. Method

3.1. Participants

36 subjects (twelve male M = 24.36; SD = 2.36), students of Psychology at the Catholic University of Milan, participated in the study. All subjects were Italians and they gave their agreement to participate at the research.

3.2. Materials

A set of sentences with literal or metaphoric meaning was used. Metaphors have to satisfy the followings characteristics:

- Each sentence was composed by 4 words, with a metaphor expressed in a nominal form (Pynte et al., 1996).
- The metaphoric content was obtained by a unique noun-term (metaphorical target), placed at the end of sentence (metaphorical ending) (Tartter et al., 2002).
- The context of each sentence was enough to activate a metaphoric decoding, without necessity of additional information.
- We tested the familiarity of metaphors (frozen metaphors), considered as salient and contextually relevant. As underlined by previous research, the degree of familiarity seems play an important role in the decoding process (Ahrens et al., 2007; Giora, 2007; Giora & Fein, 1999 [1]; Mashal et al., 2007).

We started form the metaphorical sentence to create the literal one, replacing the final metaphoric target with a literal target (a noun). So, we obtained pairs of analogous sentences in the two experimental conditions, where only the ending word was different (Tartter et al., 2002), as in the following example:
a) "the lawyers are sharks";
b) "the lawyers are professional".

The target word was balanced with respect of the mean length (metaphorical range = 2-5 syllables; literal range = 3-5 syllables). A pre-experimental phase (by 25 subjects on a 7-points Likert scale), tested familiarity (metaphor M = 6.13, SD = 0.69; literal M = 5.96, SD = 0.86) and abstractness/concreteness (metaphorical concrete M = 5.70, SD = 0.54; literal concrete M = 6.01, SD = 0.40; metaphorical abstract M = 5.89, SD = 0.67; literal abstract M = 5.81, SD = 0.47) Regarding the latter parameter, we choose the

[1] The degree of metaphor conventionality plays an important role and is a property placed along a continuum, from a minimum level (innovative and unfamiliar metaphors) to a maximum level (conventional and familiar metaphors). The idioms are set to the extreme of this potential continuum (Katz & Ferretti, 2001; Papagno, Oliveri & Romero, 2002). With conventionality, they means *stability* in our *language*, that refers to a categorial and of pre-existing conceptual system (Lakoff & Johnson, 1980; Pynte et al., 1996). For example, "love is a trip" is a conventional metaphor that belong to our conceptual system, sharing in common knowledge, that it is understood before than a new metaphor (Blank, 1988; Gentner & Wolff, 1997; Lakoff, 1993).

same number of concrete and abstract words (both literal and metaphoric), with a 50% for each category.

Moreover, we balanced the sentences in relation to the content variable: congruence (coherence of content) vs incongruence (incoherence of content) (Balconi & Pozzoli, 2004; 2005). Subsequently we created two sequences of new sentences, one with a congruent content, the other with an incongruous content, as in the following example: for literal "the soldiers are fighters", for metaphorical "the soldiers are lions". Altogether, we have created four different batteries, each containing only one version of sentence of the four possible.

Every subject was submitted to one of the four sequences, composed by 160 sentences, 80 metaphoric (equally subdivided in congruent and incongruent) and 80 literal (the same as previous). Each sequence was opportunely randomized, taking into account to not set similar sentences (for metaphorical meaning or for content) one after the other.

To control the relevance of the metaphoric/literal conditions, 25 subjects evaluated each sentence in a pre-experimental test (7-points Likert scale). For the composition of the final sequences we considered:

- as metaphorical the sentences with an evaluation of metaphoricity of M = 5 or upper (M = 6.43, SD = .38);
- as literal the sentences with an opposite evaluation, that is for metaphoricity M = 2 or lower (M = 1.90, SD = .41).

3.2. Procedure

The experiment took place in a room opportunely darkened and soundproofed. Thanks to the apparatus of stimulation (STIM 2.2) the sentences were presented at the centre of the screen, one word at a time (3 cm of height, white on black background), for a mean duration of 300 ms (SOA 600 ms) and an inter-sentence interval of 1200 ms.

Subjects were seated in front of a monitor to 100 cm of distance (visual horizontal angle of 4° and vertical of 6°), and they were asked only to read and understand the sentence, pressing the right button of the mouse when they have finished to comprehend it. A fixation point was present at the center of the screen, before that each stimulus appeared. With the aim to enable a correct familiarization with the experiment, the subjects were submitted to a pre-experimental phase (20 trials –10 metaphoric and 10 literal).

4. Data analysis and results

4.1. ERPS data

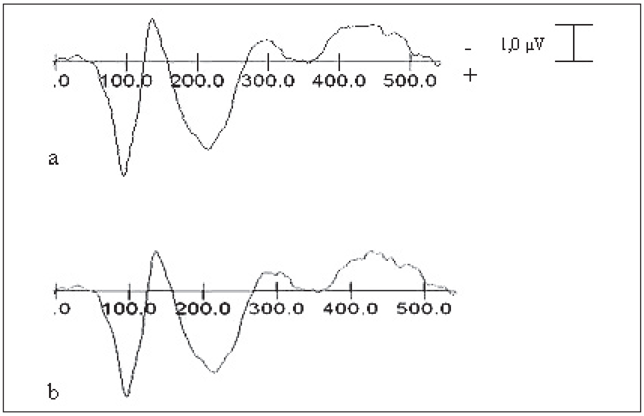
a) Recording parameters
The EEG was recorded with a 64-channel DC amplifer *(SYNAMPS* system) and acquisition software (NEUROSCAN 4.2) at 32 electrodes (International 10-20 system, Jasper, 1958) with reference electrodes at the mastoids, and mounted in a stretch-lycra electro-cup (high density registration). Electroculograms (EOG) were recorded from electrodes lateral and superior to the left eye. The signal (sampled at 256 Hz) was amplified and processed with a pass-band from .01 to 50 (off-line) Hz and was recorded in continuous mode. Impedance was controlled and maintained below 5 KΩ. Fourteen of the registered sites were considered for the statistical analysis (four central, Fz, Cz, Pz, Oz; ten lateral, F3, F4, C3, C4, T3, T4, P3, P4, O1, O2). An averaged waveform (off-line) was obtained (trials exceeding 50 μV in amplitude were excluded from the averaging process) for each type of condition (literal vs metaphoric) and content (congruent/incongruent). The EEG signals were visually scored on a high-resolution computer monitor and portions of the data that contained eye movements, muscle movements, or other source of artifact were removed. The percentage of the rejected epochs was low (6%). Peak amplitude measurement was quantified relative to 100 ms pre-stimulus (epoch duration: -100/900 ms).

b) Morphological analysis of wave profile
In order to individuate the main variations in the wave profile, we applied a qualitative, morphological analysis to the ERPs for both the literal and metaphoric conditions. As showed by the following figure (Fig. 1), in both cases it was possible to observe a previous positive peak of high amplitude around 200 ms post-stimulus (P2), and a later complex of two negative deflections, peaked at about 300 and 400 ms post-stimulus.

Then, for the statistical analysis (quantitative analysis), we considered only the two negative deflections, N300 and N400 (within the time interval of 200-400 ms and 400-600 ms).

c) Quantitative analysis of ERPs
We considered two dependent measures: the amplitude and the latency of the peaks for each deflection. An ANOVA for repeated measures was applied with three within-subjects factors: condition (2, metaphorical/literal), content (2, congruent/incongruent), electrodes (14). A Greenhouse-Geiser cor-

rection was applied in case of more than one degree of freedom. Besides the correspondent value of η^2 was calculated. In order to simply the data presentation, we reported only what was significant to the statistical analysis. With the aim to better study the localization effect, two new independent factors were calculated: one related to the four electrodes of the median line (Fz, Cz, Pz, Oz, from now the median) and the second one related to the two hemispheres (for the right one, a mean value was calculated on the electrodes F3, C3, T3, P3, O1; for the left one on the F4, C4, T4, P4, O2, from now lateralization). The variability of electrode profile was monitored in such way (Luck, 2005).

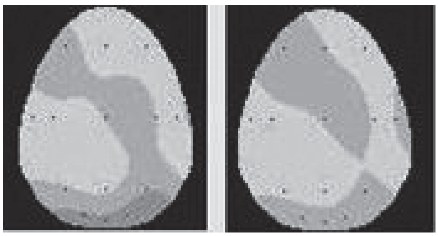


Fiure 1. Grandaverage (all the electrodes) for metaphoric (a) and literal (b) sentences

4.1.1. N3 Effect: Intensity and latency of the peak

The ANOVA for repeated measures (2 × 2 × 14) showed a significant effect for condition ($F(1,35) = 6.46$, $p < .01$, $\eta^2 = .34$) and electrode ($F(13,35) = 6.70$, $p < .01$, $\eta^2 = .37$). As suggests by mean values of peak (Table 1a), the negative deflection for metaphoric condition had greater intensity in comparison to literal sentences. Besides, the significance of the interaction effect "condition x electrode" ($F(13,35) = 4.60$, $p = .01$, $\eta^2 = .28$) allows to point out that the N3 distribution is not homogeneous on the scalp. Contrarily, no relevant differences in peak amplitude were found as a function of the content: the two conditions of congruence/incongruence showed similar profiles.

The following ANOVA better defines the cortical areas involved in the elaboration: we found an interaction effect condition x median ($F(3,35) = 6.28$, $p = .01$, $\eta^2 = .33$). Particularly, the planned comparisons (analysis of the contrasts) showed a greater peak intensity for the metaphoric stimuli in the occipital area of the scalp (Oz) in comparison to the parietal (Pz) ($F(1,35) = 7.13$, $p < .01$, $\eta^2 = .38$), the central (Cz) ($F(1,35) = 6.98$, $p < .01$, $\eta^2 = .31$) and the frontal ones (Fz) ($F(1,35) = 6.61$, $p = .01$, $\eta^2 = .29$), while literal condition doesn't show significant results. On the contrary, the lateralization effect doesn't show statistically significant differences. In the following map, we present the variations of the N3 effect on the scalp.



a b

Figure 2. Cortical maps of N3 (338 msec. latency) for metaphoric (a) and literal (b) sentences

The analysis of N3 latency doesn't show meaningful differences for condition, content and electrode effects, as well as for their interactions. The peak has a mean latency of 338 ms post-stimulus.

4.1.2. N4 Effect: Intensity and latency of peak

We applied an ANOVA for repeated measures to the N400 too, that show the effect of content ($F(1,35) = 4.54$, $p = .03$, $\eta^2 = .26$), and electrode ($F(13,35) = 7.84$, $< .01$, $\eta^2 = .40$), but not of condition. The following table (Table 2a and 2b) shows the presence of a higher peak amplitude of N4 for incongruous in comparison to congruous condition. The N4 effect is not sensitive to the condition: metaphorical and literal sentences aren't significantly differentiated for the intensity of the peak. Besides, the posterior areas of the scalp (Pz particularly) seems to be more activated in comparison to the anterior and central ones.

The successive ANOVA with the median and lateralization factors underlined more specifically the anterior-central-posterior differences (median effect $F(3,35) = 6.89$, $p < .01$, $\eta^2 = .38$), additionally to the content effect.

Table 1a. Mean value of peak amplitude of N3

	Peak amplitude[a]											
	Fz		Cz		Pz		Oz		Right		Left	
N3	*M*	*Sd*	*M*	*Sd*	*M*	*Sd*	*M*	*Sd*	*M*	*Sd*	*M*	*Sd*
congruous metaphors	-0.80	0.40	-0.74	0.15	-0.53	0.11	-0.89	0.10	-0.51	0.09	-0.71	0.07
uncongruous metaphors	-0.90	0.50	-0.95	0.15	-0.76	0.12	-1.01	0.11	-0.78	0.11	-0.87	0.09
congruous literals	-0.40	0.10	-0.48	0.13	-0.32	0.11	-0.82	0.13	-0.36	0.11	-0.63	0.09
uncongruous literals	-0.50	1.20	-0.70	0.12	-0.48	0.13	-0.95	0.12	-0.58	0.10	-0.73	0.07

[a] in μvolt.

Table 1b. Mean value of peak latency for N3

	Peak latency[b]											
	Fz		Cz		Pz		Oz		Right		Left	
N3	*M*	*Sd*	*M*	*Sd*	*M*	*Sd*	*M*	*Sd*	*M*	*Sd*	*M*	*Sd*
congruous metaphors	346.12	3.70	340.19	4.10	342.58	3.86	337.86	3.72	341.38	2.70	336.94	3.19
uncongruous metaphors	340.92	2.34	338.47	4.43	338.58	5.13	343.58	3.58	342.17	3.04	335.89	3.43
congruous literals	336.53	2.89	338.89	4.97	343.06	4.28	341.69	2.93	339.97	2.83	334.06	2.85
uncongruous literals	341.23	3.21	331.72	5.17	338.36	4.79	342.58	3.66	339.61	3.23	335.96	3.09

[b] in msec.

Table 2a. Mean value of peak amplitude of N4

	Peak amplitude											
	Fz		Cz		Pz		Oz		Right		Left	
N4	*M*	*Sd*	*M*	*Sd*	*M*	*Sd*	*M*	*Sd*	*M*	*Sd*	*M*	*Sd*
congruous metaphors	-1.30	0.13	-1.24	0.13	-1.24	0.13	-1.22	0.09	-0.98	0.70	-1.11	0.08
uncongruous metaphors	-1.20	0.20	-1.21	0.12	-1.11	0.11	-1.26	0.09	-1.11	0.10	-1.13	0.10
congruous literals	-1.10	0.13	-1.03	0.13	-1.30	0.10	-1.13	0.08	-0.86	0.80	-1.01	0.08
uncongruous literals	-1.40	0.13	-1.13	0.13	-2.00	0.12	-1.31	0.10	-1.50	0.10	-1.11	0.08

Table 2b. Mean value of peak latency for N4

	Peak latency											
	Fz		Cz		Pz		Oz		Right		Left	
N4	*M*	*Sd*	*M*	*Sd*	*M*	*Sd*	*M*	*Sd*	*M*	*Sd*	*M*	*Sd*
congruous metaphors	432.43	4.32	425.42	5.35	425.47	4.89	426.92	4.60	427.69	3.44	430.25	3.19
uncongruous metaphors	439.32	4.89	426.83	5.63	430.50	5.36	428.78	4.32	429.62	3.82	427.78	3.28
congruous literals	431.21	3.23	429.78	6.01	431.06	5.97	423.94	5.42	431.58	3.93	431.03	4.56
uncongruous literals	430.28	5.04	428.81	6.23	439.67	6.18	429.92	4.91	433.76	4.13	433.68	4.29

The contrast analysis showed a greater contribution of Pz in comparison with the other locations (Fz ($F(1,35) = 6.41$, $p < .01$, $\eta^2 = .34$); Cz ($F(1,35) = 5.39$, $p < .01$, $\eta^2 = .28$); Oz ($F(1,35) = 4.48$, $p < .01$, $\eta^2 = .26$). The significance of the interaction effect content x lateralization ($F(1,35) = 7.09$, $p < .01$, $\eta^2 = .42$) and of content x median ($F(3,35) = 6.80$, $p < .01$, $\eta^2 = .35$) revealed the presence of a greater peak amplitude of N4 for Pz in the condition of incongruence in comparison to all the other electrodes (respectively Fz ($F(1,35) = 8.04$, $p < .01$, $\eta^2 = .44$); Cz ($F(1,35) = 6.12$, $p < .01$, $\eta^2 = .32$); Oz ($F(1,35) = 6.78$, $p < .01$, $\eta^2 = .40$). Then, it was shown a greater intensity of peak for incongruous sentences in the right areas in comparison to left ones ($F(1,35) = 8.54$, $p < .01$, .2), while significant differences were not observed for congruent condition.

Significant results were not found for the latency dependent measure.

4.1.3. Behavioural index (RT)

Behavioral data of 32 subjects were used for the statistical analysis, since 4 subjects were eliminated for their too many missing responses (more than 10%). The ANOVA for repeated measures, applied to condition (2) and content (2) independent factor, didn't reveal significant differences. A particular data, even without a significant inferential statistic, is to underline: metaphoric sentences registered anticipated RT values in comparison to literal condition (specifically for metaphoric sentences M = 224.64 ms; SD = 2.45; for literal sentences M = 237.75 ms; SD = 3.45).

5. Discussion

The data we found allowed us to investigate the cognitive correlates in understanding metaphors. We synthesize the main results into the following main points.

1. First of all, N4 effect didn't appear as a marker of the elaboration of metaphoric meaning, but it is mostly sensitive to the content variations (congruence/incongruence) of the sentences.
2. The absence of interaction effect between condition and content suggests that the incidence of the content on modulation of N4 appears entirely independent from the presence of metaphoric or literal meaning decoding process. It seems to be localizated in the parietal areas of the scalp (PZ) for the condition of incongruence.

3. The second ERP effect considered, the N3 deflection, appears sensitive only to the condition effect: it was mostly present for the elaboration of metaphoric meanings. The absence of significance in relation to factor content, suggests that it hasn't a direct relationship with the presence of semantic anomalies in the stimuli. Additionally, its cortical distribution involves mainly the posterior areas of the scalp (Oz), above all for metaphoric condition.
4. Lateralization effect (right vs left hemisphere) doesn't show statistical significance in N3 peak amplitude, while the N4 index appears to mostly activate the cortical right areas in condition of content incongruity.
5. Behavioral data show the absence of meaningful differences, and in particular in relationship to the dichotomy metaphorical/literal.

At first, the results about the N4 ERP effect are relevant in the debate about the relationship between metaphoric and literal decoding process. The absence of significant differences in the two conditions doesn't allow to consider the figurative language comprehension after the inhibition of the standard, literal process. These results are in line with the direct access model, for which metaphor appears cognitively and semantically not anomalous at the semantic level.

The approach of some cognitive models (Coulson & Van Petten, 2002), for which metaphor requires a more complex process of elaboration, doesn't find confirm here, and response times too show that there is not a deceleration in understanding metaphor, but it is automatically and immediately activated as literal decoding. Behavioural data about literal/metaphorical condition give even more interesting information: metaphorical comprehension produces an observable reduction of RT in comparison with literal comprehension. Such data could be explained considering the high value of familiarity of our metaphoric sentences: the metaphors we used are conventional, cognitively salient for the subject. As suggested by Giora's model (Giora & Fein, 1999), the salience and the familiarity are reasonable factors for the metaphoric decoding: high-salient metaphors represent the automatic option activated.

The latency of the N4 effect doesn't show differences in relationship to the experimental variables: this tendency is important especially for the condition factor, confirming the homogeneity of metaphorical and literal decoding.

Besides, we confirmed results of other studies, detecting a negative deflection of N4 sensible to the semantic congruence of the stimulus: the significant effect produced by the congruence/incongruence of sentence (and not by the metaphorical/literal condition) would confirm that metaphoric comprehension does not require a decoding process qualitatively different from the literal one. On the whole, we can suppose the existence of a kind of continuum between literal, metaphoric conventional and metaphoric unconventional context. This intermediate position of metaphoric conventional would be due to the membership of familiar metaphorical meaning to a pre-existing semantic system, which makes them "conventional".

A second main issue here examined is referred to the different wave profiles of N3 effect for metaphoric sentences in comparison to literal condition, while the first registered a meaningful higher than the second one. This result, associated to the absence of N4 modifications, would rather suggest an access to a different *representation store* for the metaphor meanings, whose nature might be iconic (Balconi & Tutino, 2006).

Iconic elaboration process could be therefore involved in a significant measure, even though not exclusive. Previous researches found N3 effect for image-based representations, or representative of a pictorial code, in response to iconic stimuli (such as script compared to lexical stimuli) (Holcomb et al., 1999; Paivio, 1991; Shallice, 1993; West & Holcomb, 2002). The N3 localization, that is significantly more present in occipital area of the scalp, would confirm the pictorial nature of the negative deflection. Neuropsychological studies have postulated, in fact, that such areas are extensively responsible for the elaboration of visual stimuli (primary visual areas) and they are mainly involved in processes that implicate an elaboration mediated by the images (Ladavas & Berti, 1995).

On the contrary, we didn't found the expected lateralization effect in metaphor decoding.

The researches about hemispheric asymmetry in language elaboration, in fact, postulated the right hemisphere as mostly responsible of the pragmatic components of language (Beeman & Chiarello, 1998) and particularly of metaphoric decoding (Gineste et al., 2000; Mashal, Faust & Hendler, 2005; Mashal et al., 2007). Contrarily, the left hemisphere was considered more implicated in the lexical processes and, in general, for the elaboration of the standard or literal meaning. Nevertheless, the high degree of familiarity of our metaphoric stimuli could have effect in the similar contribution of both the hemispheres, with a joined action of the right and left areas of the scalp (Schmidt et al., 2007; Stringaris et al., 2007).

Moreover, the hemispheric differentiation appeared to be sensitive to sentence congruousness, with a greater right activation for semantically anomalous stimuli and a substantial homogeneity of two hemispheres for congruent utterances.

This particular finding could be explained with the hypothesis that the right hemisphere in general would clear the ambiguous meaning, or it would operate to choose an alternative meaning to that anomalous. Coney and

Evans (2000) proposed the paradigm of the lexical ambiguity to explain the different contribution of the two hemispheres in language decoding, where the relationship between context and sentence meaning plays a fundamental role. For the authors, the left hemisphere would be preferably activated when exist a concordance between the contextual domain and the dominant meaning, while the right hemisphere give a greater contribution for not-dominant or anomalous meaning. According with this point of view, the left hemisphere would be more selective, choosing an unique semantic option between those possible (Faust & Lavidor, 2003), and the right hemisphere has the function to support it in order to make accessible a greater ensemble of alternative and less probable meanings (Jung-Beeman, 2005).

In synthesis, we found the presence of a negative ERPs complex composed by two different ERP indexes: the first (N300) is presumably a marker of the "iconic" format of representation, and the second (N400) is involved in the semantic elaboration of the stimuli (congruousness effect), testing the consonance between sentence-meaning and sentence-context. Nevertheless, our results show a substantial difference in the processes involved in metaphorical and literal elaboration, based on the contribution of different mechanisms that make specific and unique metaphoric meaning representation compared with literal meaning: the iconic format of representation is the main point that must be considered in order to explain the main differences with the literal format of representation.

Moreover, these results are opposite with theory of linguistic anomaly or indirect access, indicating that the difference is not for the presence of differences in terms of degree of complexity in sentence elaboration or of anomalous/not anomalous meaning comprehension.

In general, our results would suggest a direct resemblance between the semantic comprehension of metaphors and literal sentences decoding, even if it involves different representational format. Nevertheless, the high level of conventionalization of metaphor may have facilitate the comprehension of the sentence meaning, making more easier to compare metaphoric with literal decoding process. Moreover, the balanced contribution of concrete/abstract target in the sentences construction is a fact to be evaluated, since this feature was shown to be directly related to the level of imaginability (West & Holcomb, 2002). An explicit comparison between these two categories of target stimuli may produce some interesting effect on the ERP modulation, specifically on the N3 negative deflection.

References

Ahrens, K., Liu, H.-L., Lee, C.-Y., Gong, S.-P., Fang, S.-Y., & Hsu, Y.-Y. (2007). Functional MRI of conventional and anomalous metaphors in Mandarin Chinese. *Brain and Language*, 100, 163-171.

Balconi, M., & Lucchiari, C. (2005). Consciousness, emotion and face: An event-related potential study. In: Ellis, R., Newton, R.N. (eds.), *Consciousness and Emotion. Agency, Conscious Choice, and Selective Perception*. Philadelphia: Benjamins, pp. 121-135.

Balconi, M., & Pozzoli, U. (2003). Face-selective processing and the effect of pleasant and unpleasant emotional expression on ERP correlate. *International Journal of Psychophysiology*, 49, 67-74.

Balconi, M., & Pozzoli, U. (2004). Elaborazione di anomalie semantiche e sintattiche con stimolazione visiva e uditiva. Un'analisi mediante correlati ERPs. *Giornale Italiano di Psicologia*, 3, 585-612.

Balconi, M., & Pozzoli, U. (2005). Comprehending semantic and grammatical violations in Italian. N400 and P600 comparison with visual and auditory stimuli. *Journal of Psycholinguistic Research*, 34, 71-98.

Balconi, M. & Tutino, S. (2006). A fighter is a lion. Neuropsychological and cognitive processes in decoding a metaphor. An analysis through ERPs. *Journal of the International Neuropsychological Society, Supplement*, 12, 88.

Beeman, M., & Chiarello, C. (1998). Concluding remarks: Getting the whole story right. In: Beeman, M., Chiarello, C. (eds.), *Right Hemisphere Language Comprehension: Perspectives from Cognitive Neuroscience*. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 377-389.

Blank, G.D. (1988). Metaphors in the lexicon. *Metaphor and Symbolic Activity*, 3, 21-36.

Bonnaud, V., Gil, R., & Ingrand, P. (2002). Metaphorical and non-metaphorical links: A behavioral and ERP study in young and elderly adults. *Clinical Neurophysiology*, 32, 258-268.

Bottini, G., Corcoran, R., Sterzi, R., Paulesu, E., Schenone, P., Scarpa, P., Frackowiak, R.S.J., & Frith, C.D. (1994). The role of the right hemisphere in the interpretation of figurative aspects of language. A positron emission tomography activation study. *Brain*, 117, 1241-1253.

Coney, J., & Evans, K.D. (2000). Hemispheric asymmetries in the resolution of lexical ambiguity. *Neuropsychologia*, 38, 272-282.

Coulson, S., & Van Petten, C. (2002). Conceptual integration and metaphor: An event-related potential study. *Memory and Cogntion*, 30, 958-968.

Eddy, M., Schmid, A., & Holcomb, P.J. (2006). Masked repetition priming and event-related brain potentials: A new approach for tracking the time-course of object perception. *Psychophysiology*, 43, 564-568.

Faust, M., & Lavidor, M. (2003). Semantically convergent and semantically divergent priming in the cerebral hemispheres: Lexical decision and semantic judgment. *Cognitive Brain Research*, 17, 585-597.

Federmeier, K.D., & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 202-224.

Gentner, D., & Wolff, P. (1997). Alignment in the processing of metaphor. *Journal of Memory and Language*, 37, 331-355.

Gibbs, R. (1994). *Poetics of Mind: Figurative Thought, Language, and Understanding.* Cambridge, MA: Cambridge University Press.

Gibbs, R.W., & Gerring, R.J. (1989). How context makes metaphor comprehension seem "special". *Metaphor and Symbolic Activity*, 4, 145-158.

Gineste, M., Indurkhya, B., & Scart, V. (2000). Emergence of Features in Metaphor Comprehension. *Metaphor and Symbol*, 15, 117-135.

Giora, R. (2003). *On our Mind: Salience, Context, and Figurative Language*. New York: Oxford University Press.

Giora, R. (2007). Is metaphor special? *Brain and Language*, 100, 111-114.

Giora, R., & Fein, O. (1999). On understanding familiar and less-familiar figurative language. *Journal of Pragmatics*, 31, 1601-1618.

Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends in Cognitive Science*, 7, 92-96.

Grady, J. (2005). Primary metaphors as inputs to conceptual integration. *Journal of Pragmatics. Special Issue: Conceptual Blending Theory*, 37, 1595-1614.

Grice, H.P. (1975). Logic and Conversation. In: Cole, P., Morgan, J. (eds.), *Syntax and Semantics*, vol. 3. New York: Academic Press, pp. 41-58.

Hamm, J.P., Johnson, B.W., & Kirk, I.J. (2002). Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. *Clinical Neurophysiology*, 113, 1339-1350.

Handy, T.C., & Khoe, W. (2005). Attention and sensory gain control: A peripheral visual process? *Journal of Cognitive Neuroscience*, 17, 1936-1949.

Holcomb, P.J., Kounios, J., Anderson, J.E., & West, W.C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology. Learning Memory and Cognition*, 25, 721-742.

Iakimova, G., Passerieux, C., Laurent, J., & Hardy-Bayle, M. (2005). ERPs of metaphoric, literal, and incongruous semantic processing in schizophrenia. *Psychophysiology*, 42, 380-390.

Jasper, H.H. (1958). The ten-twenty electrode system of International Federation EEG. *Clinical Neurophysiology*, 10, 371-375.

Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. *Trends in Cognitive Science*, 9, 512-518.

Kacinik, N.A., & Chiarello, C. (2007). Understanding metaphors: Is the right hemisphere uniquely involved? *Brain and Language*, 100, 188-207.

Katz, A.N., & Ferretti, T.R. (2001). Moment-by-moment readings of proverbs in literal and nonliteral contexts. *Metaphor and Symbol*, 16, 193-221.

Kazmerski, V.A., Blasko, D.G., & Dessalegn, B.G. (2003). ERP and behavioral evidence of individual differences in metaphor comprehension. *Memory and Cognition*, 31, 673-689.

Kintsch, W. (1988). The role of knowledge in discourse comprehension construction-integration model. *Psychological Review*, 95, 163-182.

Kutas, M., & Hillyard, S.A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203-205.

Ladavas, E., & Berti, A. (1995). *Neuropsicologia*. Bologna: Il Mulino.

Laeng, B., Zarrinpar, A., & Kosslyn, S.M. (2003). Do separate processes identify objects as exemplars versus members of basic-level categories? Evidence from hemispheric specialization. *Brain and Cognition*, 53, 15-27.

Lakoff, G. (1993). The contemporary theory of metaphor. In: Ortony, A. (ed.), *Metaphor and Thought*, 2nd. ed. Cambridge, MA: Cambridge University Press, pp. 202-251.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live by.* Chicago: University of Chicago Press.

Large, M.E., Kiss, I., & McMullen, P.A. (2004). Electrophysiological correlates of object categorization: back to basics. *Cognitive Brain Research*, 20, 415-426.

Luck, S.J. (2005). Ten simple rules for designing ERP experiments. In: Handy, T.C. (ed.), *Event-related Potentials. A Method Handbook*. Cambridge, MA: The MIT Press, pp. 17-33.

Mashal, N., Faust, M., & Hendler, T. (2005). The role of the right hemisphere in processing nonsalient metaphorical meanings: Application of Principal Components Analysis to fMRI data. *Neuropsychologia*, 43, 2084-2100.

Mashal, N., Faust, M., Hendler, T., & Jung-Beeman, M. (2007). An fMRI investigation of the neural correlates underlying the processing of novel metaphoric expressions. *Brain and Language*, 100, 115-126.

McElree, B., & Nordlie, J. (1999). Literal and figurative interpretations are computed in equal time. *Psychonomic Bulletin & Review*, 6, 486-494.

McGlone, M.S., & Manfredi, D.A. (2001). Topic-vehicle interaction in metaphor comprehension. *Memory & Cognition*, 29, 1209-1219.

Newman, S.D., Just, M.A., & Mason, R.A. (2003). Understanding text with the right side of the brain: What functional neuroimaging has to say. In: Tomitch, L.M.B., Rodrigues, C. (eds.), *Ensaios Sobre a Linguagem e o Cérebro Humano: Contribuições Multidisciplinares*, pp. 71-84.

Ortony, A. (1979). *Metaphor and Thought*. Cambridge, MA: Cambridge University Press.

Paivio, A. (1991). *Images in Mind: The Evolution of a Theory.* New York: Harvester Wheatsheaf.

Papagno, C., & Caporali, A. (2007) submitted. Testing idiom comprehension in aphasic patients: The modality and the type of idiom effects. *Brain and Language*.

Papagno, C., Oliveri, M., & Romero, L. (2002). Neural correlates of idiom comprehension. *Cortex*, 38, 895-898.

Pynte, J., Besson, M., Robichon., F.H., & Poli, J. (1996). The time-course of metaphor comprehension: An event-related potential study. *Brain and Language*, 55, 293-316.

Rugg, M.D., & Coles, M.G.H. (1995). The ERP and cognitive psychology: Conceptual issues. In: Rugg, M., Coles, M. (eds.), *Electrophysiology of Mind*. Oxford: Oxford University Press.

Schmidt, G., DeBuse, C., & Seger, C. (2007). Right hemisphere metaphor processing? Characterizing the lateralization of semantic processes. *Brain and Language*, 100, 127-141.

Searle, J.R. (1979). *Speech Acts.* Cambridge: Cambridge University Press.

Shallice, T. (1993). Multiple semantics: Whose confusions? *Cognitive Neuropsychology*, 10, 251-261.

Stringaris, A., Medford, N., Giora, R., Giampietro, C.V., Brammer, J.M., & David, S.A. (2006). How metaphors influence semantic relatedness judgments: The role of the right frontal cortex. *Neuro Image*, 33, 784-793.

Tartter, V.C., Gomes, H., Dubrovsky, B., Molholm, S., & Stewart, R.V. (2002). Novel metaphors appear anomalous at least momentarily: Evidence from N400. *Brain and Language*, 80, 488-509.

West, W.C., & Holcomb, P.J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research*, 13, 363-375.

Wolff, P., & Gentner, D. (2000). Evidence for role-neutral initial processing of metaphors. *Journal of Experimental Psychology. Learning Memory and Cognition*, 26, 529-554.

A validated battery of vocal emotional expressions

Pierre Maurage [1] **- Frédéric Joassin** [1] **- Pierre Philippot** [1]
Salvatore Campanella [2]

[1] *Cognitive Neurosciences and Clinical Psychology Research units, Department of Psychology, Catholic University of Louvain, Louvain-la-Neuve, Belgium*

[2] *Department of Psychiatry, Brugmann Hospital, Free University of Brussels, Brussels, Belgium*

pierre.maurage@psp.ucl.ac.be

Abstract

For a long time, the exploration of emotions focused on facial expression, and vocal expression of emotion has only recently received interest. However, no validated battery of emotional vocal expressions has been published and made available to the researchers' community. This paper aims at validating and proposing such material. 20 actors (10 men) recorded sounds (words and interjections) expressing six basic emotions (anger, disgust, fear, happiness, neutral and sadness). These stimuli were then submitted to a double validation phase: (1) preselection by experts; (2) quantitative and qualitative validation by 70 participants. 195 stimuli were selected for the final battery, each one depicting a precise emotion. The ratings provide a complete measure of intensity and specificity for each stimulus. This paper provides, to our knowledge, the first validated, freely available and highly standardized battery of emotional vocal expressions (words and intonations). This battery could constitute an interesting tool for the exploration of prosody processing among normal and pathological populations, in neuropsychology as well as psychiatry. Further works are nevertheless needed to complement the present material.

Keywords: Emotional Vocal Expression; Prosody; Battery

1. Introduction

Every banal or crucial moment of our existence is associated with an emotional feeling. Indeed, emotions are an essential aspect of life, they influence our behaviours, decisions, motivations and social interactions. Emotions have been explored since the early development of modern experimental psychology, and constituted a major field of investigation during the last decades. Nevertheless, the debate concerning the emotional phenomena is still brisk. Many models and interpretations have been proposed in social psychology (e.g. Leventhal, 1984), cognitive psychology (e.g. Frijda et al., 1989), and more recently in a neuroscience perspective (Davidson, 2003). Despite this theoretical disparity, a consensus progressively emerged concerning the global definition of emotion: A multidimensional concept (e.g. Scherer, 1984) including at least some physiological changes, an internal state (namely an "affect"), a cognitive behaviour, but also an overt behaviour.

Behaviours constitute a central aspect of emotion, and the behavioural correlates associated with every emotion thus became a central field of research. The expression of the emotional state and its corollary, the perception of emotions are crucial for the social life: emotions are communicative (Oatley & Johnson-Laird, 1987). Among the various modes of emotional expression (e.g. tears, cries, physical aggression, …), the face appears to be a central medium for communication, and it rapidly became the central focus of researchers. It has notably been shown that each basic emotion (mainly fear, anger, happiness, sadness, disgust and surprise) is associated with a specific activation pattern of the face's muscles (as described in the FACS, Ekman & Friesen, 1976) and that most of these facial expressions are expressed and recognized universally, regardless of the culture (Ekman et al., 1969; Izard, 1977). This research allowed the construction of standardized batteries of EFE (emotional facial expressions, e.g. Ekman & Friesen, 1976; Hess & Blairy, 1995; Matsumoto & Ekman, 1988). With the support of these validated stimuli, hundreds of studies related to the expression and recognition of emotional facial expression were carried out in various fields, such as social psychology (e.g. Ackerman et al., 2006), cognitive psychology (e.g. Campanella et al., 2002), neuroscience (e.g. Blair et al., 1999) or psychopathology (e.g. Minzenberg et al., 2006). Thus, the construction of reliable materials allowed the fast expansion of the studies exploring the EFE and facilitated the comparison among studies.

More recently, notably with the development of crossmodal explorations (e.g. Calvert, 2001), a growing interest has been directed towards other sensory modalities. Concerning emotional processing, studies have been carried out to explore the perception of olfactory (e.g. Hermans et al., 2005), tactile (Montoya & Sitges, 2006) and gustatory (Greimel et al., 2006) stimuli, but the most flourishing domain concerns the auditory stimulations. Indeed, the emotional valence conveyed by sounds recently gave rise to various studies, mainly exploring prosody among normal (see Ethofer et al., 2006 for a review) and pathological populations (e.g. Monnot et al., 2002). Despite this growing interest for the auditory expression of emotion, and despite a great amount of earlier works on emotional prosody (e.g. Banse & Scherer, 1996; Kappas et al., 1991; Hess et al., 1989), no study has yet been conducted to construct a standardized, validated and freely available battery of emotional auditory words and interjections. Actually, the existing sets of stimuli are, to our knowledge, either partials (i.e. not proposing the complete range of emotions), or only partially validated and standardized (e.g. Fecteau et al., 2005; Golan et al., 2006; 2007; Pourtois et al., 2005). As a consequence, most of the studies on emotional prosody have been based on ad hoc constructed stimuli, which constitutes undoubtedly a loss of time and makes difficult the comparison among studies.

It thus appears crucial, in order to facilitate future studies and to propose a common reference freely and easily available, to construct a standardized and validated battery, following the example of what has been done concerning the EFE. This study aims at proposing an emotional auditory battery using various actors, stimuli and emotions. This validated auditory battery is made freely accessible to all researchers via the Internet.

2. Methods

2.1. Stimuli construction

2.1.1. Actors, emotions and stimuli

Twenty actors (10 men and 10 women) were recruited in Louvain-la-Neuve, Belgium. All actors were native French speakers. It was decided to record the five of the most used basic emotions (namely anger, fear, happiness, sadness and disgust), and the neutral state. Short emotional scripts were built in order to ensure that emotions were clearly defined. Before the recording of each emotion, actors read a brief text describing a situation eliciting this emotion. Two types of stimuli were chosen: One consisted in a meaningless sound (namely the interjection “ah”) and the other in a semantically neutral word (namely the French word “papier”, i.e. “paper”).

2.1.2. Design and recording

Each actor produced two times each emotion and stimulus type, leading to the recording of 480 voice samples: 10 (actors) X 2 (sex of actor) X 6 (emotions) X 2 (kinds of stimulus) X 2 (samples). The session took place at the Auditory Lab of the Faculty of Psychology in Louvain-la-Neuve, and the sounds were recorded on a high-quality microphone with Goldwave 4.25 (Goldwave Inc., 2001). Before the recording of each stimulus, actors were asked to read the emotional script associated with the emotion and to imagine experiencing the situation. Each stimulus was recorded at least two times, and more if the actor or the researcher decided that the sound quality was not satisfying. There was no time limit for the recording session. After recording, a standardization of duration was imposed with Goldwave 4.25 in order to have a common length (initially varying from 630 to 792 ms) of 700 ms for all records.

2.1.3. Expert rating

Before the beginning of the experimental validation of the battery per se, a preselection was carried out on the basis of expert rating. In order to select the best recordings among each pair (as every stimulus was recorded twice) and to eliminate obviously invalid samples, three experts (researchers chosen on the basis of their knowledge on emotional prosody) inspected the 480 recordings. The three experts heard the 480 sounds and were asked on the one hand to decide which emotion was expressed (namely to detect the emotion mainly depicted in the stimulus) and on the other hand to rate them on two Lickert scales (from 1 to 7) respectively evaluating the quality of the recording and the recognizability of the emotion. The sounds were chosen if they met the following criteria: (a) The stimulus was identified by the three experts as mainly depicting the correct emotion; (b) The mean quality rating was at least of 4; (c) The mean recognizability rating was at least of 5. On this basis, 220 sounds were selected for the experimental validation, performed by 9 female and 8 male actors.

2.2. Stimuli validation

70 undergraduate psychology students at the University of Louvain (39 women and 21 men, mean age: 19.6 years) took part in the stimuli validation phase. Six sessions were conducted, each one with 10 students. The 220 stimuli were randomly presented with professional quality speakers. Each stimulus was presented once. After each presentation, participants had to report which emotion(s) were displayed in the record, with 8 points Lickert scales (from 0 “Absolutely not” to 7 “Totally”) which had to be completed on

the basis of the following general question: “Is this emotion expressed in this sound?”. Ten scales were proposed, reflecting emotions actually displayed in several stimuli (anger, fear, sadness, happiness, disgust) or other emotions (shame, boredom, surprise, contempt, annoyance) chosen for their proximity with the emotions actually expressed.

3. Results

3.1. Preselection

A first preselection was conducted. It was decided to exclude from the final battery any stimulus that had not been correctly identified by at least 90% of the participants. An identification was considered as correct when the emotion displayed by the actor obtained the highest score on the Lickert scale as compared to other emotional scales. On this basis, 25 sounds were excluded.Brain and the final battery is thus composed of 195 stimuli, each one correctly identified as depicting the expected emotion.

3.2. Basic emotion ratings

While the correct identification of each sound could be considered as a sufficient criteria to validate the battery (e.g. Banse & Scherer, 1996), this study aimed at conducting a sharper analysis. We computed the overall mean of the ratings for each stimulus on each scale. This procedure first allowed to determine the intensity of the emotion rating (namely the mean score for the expected emotion). Nevertheless, a high intensity score is insufficient to conclude that a stimulus is valid, as it could be that a stimulus is highly rated on the depicted emotion, but also on scales linked to irrelevant emotions. In order to take into account the difference between the relevant score and the other scores,Brain anda second factor was computed, namely the specificity of the emotion rating (i.e. the percentage stemming from the mean score for the expected emotion divided by the mean global scoring for the five scales associated with the basic emotions). Concerning the neutral stimulus, a sound was considered as neutral when the mean rating was lower than 1.5 for each scale associated with a basic emotion (anger, disgust, fear, happiness, sadness). On this basis we obtained, for each emotion, a set of stimuli depicting this emotion: 32 anger stimuli, 30 disgust stimuli, 34 fear stimuli, 27 happiness stimuli, 38 neutral stimuli and 34 sadness stimuli A ranking of the stimuli for each emotional category was then carried out. The intensity and specificity of the ten best sounds for each emotional category are reported in Table 1. These intensity (higher than 4.9) and specificity (higher than 75 percent) appear satisfying, with very low confusion between the six basic emotions. Moreover, Table 2 presents for each emotional set of stimuli, the mean intensity values obtained on each emotion scale (namely the emotion rating for the depicted emotion but also for the other irrelevant emotions). This Table thus shows the distinctive emotional features for each set, and the specificity of the stimuli: high ratings for the depicted emotion (mean intensity higher than 3.3) and low ratings for the other emotion scales (mean intensity lower than 0.9).

Finally, two complementary measures were computed in order to confirm the (1) the specificity of each emotional set as compared to other ones and (2) the internal consistency of each emotional set. First, two-tailed Pearson's correlations have been computed between the mean ratings for each emotion scale among the different emotion sets. As expected, no significant correlations have been detected ($p > .40$ for every correlation), showing a high independence between the ratings in the different emotion sets, and thus the specificity of each emotion set. Second, Cronbach's alphas have been computed to test the internal consistency of each emotional set on each emotion scale across participants. High values ($\alpha > 0.82$) were found for the relevant emotion scale of each emotion set (e.g. the emotion scale of anger for the anger set of stimuli), but also for the irrelevant ones ($\alpha > 0.74$), thus reinforcing the validity of the battery.

Table 1. Intensity and specificity of the ten best stimuli for each basic emotion [mean values (SD)]

Emotion type	Mean intensity[a]	Mean specificity[b]
Anger	5.44 (0.25)	83.2 (8.87)
Disgust	5.45 (0.41)	88.6 (7.08)
Fear	5.59 (0.20)	75.1 (6.97)
Happiness	5.36 (0.36)	97.6 (2.36)
Sadness	4,86 (0.68)	76.3 (9.92)
Neutral	0,09 (0.04)	/

[a] Mean score for the expected emotion, from 0 (emotion absolutely not depicted in the stimulus) to 7 (emotion fully depicted in the stimulus).

[b] Mean percentage stemming from the mean score for the expected emotion divided by the mean global scoring for the five scales associated with the basic emotions.

Table 2. Mean intensity ratings on each emotional scale for each emotion set [mean values (SD)]

Emotion Scale Emotional Set	Anger	Disgust	Fear	Happiness	Sadness
Anger (32[a])	**4.26 (1.20)**[b]	0.89 (0.58)	0.26 (0.31)	0.08 (0.29)	0.18 (0.21)
Disgust (30)	0.36 (0.40)	**3.60 (1.54)**	0.36 (0.37)	0.37 (0.48)	0.23 (0.32)
Fear (34)	0.17 (0.22)	0.89 (0.68)	**4.61 (0.91)**	0.12 (0.28)	0.88 (0.73)
Happiness (27)	0.04 (0.07)	0.39 (0.55)	0.24 (0.41)	**3.78 (1.40)**	0.12 (0.20)
Sadness (34)	0.10 (0.15)	0.48 (0.49)	0.74 (0.59)	0.13 (0.24)	**3.33 (1.25)**
Neutral (38)	0.13 (0.17)	0.18 (0.21)	0.09 (0.14)	0.14 (0.25)	0.47 (0.34)

[a] Number of stimuli in this emotional set.

[b] The score for the emotion scale associated with the depicted emotion is presented in bold type.

3.3. Secondary emotion ratings

Moreover, we explored the most frequent confusions in the emotional rating. Indeed, if there were very few misinterpretations among the six emotions depicted by the stimuli (as shown above), which was the central aim of this study, it was still possible to find more mistakes in the ratings associated with other emotions (namely shame, boredom, surprise, contempt and annoyance), as these secondary emotions are known to be often confused with the basic ones. Nevertheless, we found globally low ratings for these secondary emotions: Global mean rating was lower than 0.8 on each secondary emotion scale, and no stimulus was rated higher on a secondary scale than on the correct scale (i.e. the scale associated with the emotion actually depicted). This confirmed the specificity of the stimuli. Actually, the more frequent errors were, as expected, to confound (1) annoyance with anger (annoyance mean rating for anger stimuli was 2.9), (2) contempt with disgust (mean: 1.2), (3) surprise with fear (mean: 1.4) or happiness (mean: 1.4), (4) boredom with neutral (mean: 1.8) or sadness (mean: 1.2). However, these erroneous ratings stay relatively low, and all the other mean ratings for the secondary emotion are lower than 1, which confirms the validity of this battery.

The complete results are presented in Annex 1 that gives an exhaustive description of each stimulus: number, emotion depicted, type (word or interjection), actor, mean rating for each emotional scale, ranking and specificity percentage for each emotion.

4. Discussion

The aim of this study was to design a validated battery of emotional auditory stimuli. Indeed, the exploration of the behavioural, electrophysiological and neuroanatomical (e.g. Everhart et al., 2006; Kotz et al., 2006) correlates of emotional sound processing became a central field of interest during the last decade. Nevertheless, while a wide range of reliable materials exist for the visual emotional stimuli (e.g. Ekman & Friesen, 1976; Hess & Blairy, 1995) and while some preliminary works have been conducted to validate prosody batteries (Borod et al., 1990; Pell, 2002), only some partial or specific materials exist. For example, Fecteau et al. (2005) only used non-linguistic vocalizations; Pourtois et al. (2005) proposed only two emotions (i.e. fear and happiness); Golan et al. (2006; 2007), in the Cambridge Mindreading Face-Voice Battery, presented a battery based on emotional sentences, specifically dedicated to autistic populations and focusing on complex emotions; finally Banse and Scherer (1996) only described a partial validation of their results, with low recognition rates for certain emotions (e.g. 14% for disgust). Moreover, in most of the studies using emotional auditory stimuli, the first aim was not to present an auditory battery per se, but rather to explore various processing on the basis of auditory stimuli. These studies were thus not precisely describing the characteristics (validation, standardization, …) of the stimuli used, and did not put their stimuli at the researchers' community's disposal. As a consequence, to our knowledge, no complete, validated and free auditory emotional battery has been published up to now. This surprising absence of available material dampens the development of the field (as every researcher has to build and validate his/her own stimuli), and hamper the comparison among studies (as the specificity and intensity of the auditory emotions highly varies across experiments). This paper is thus a first attempt to develop a validated and standardized auditory battery of emotional sounds.

The advantages of this battery are the following: First, it proposes a complete and various set of stimuli, as the final battery presents 195 stimuli depicting the six basic emotions (anger, disgust, fear, happiness, neutral and sadness), with two types of stimuli (word and interjection) and a wide range of voices (eight male and nine female actors). Second, the 195 stimuli have standardized duration (700ms), which is crucial in studies where the timing of stimulation is important (e.g. studies based on reaction times recording, electrophysiological and neuroimagery studies). Third, the battery is highly validated, as the large qualitative validation (70 participants) with 10 emotional scales led to a precise ranking, based on specificity and intensity of each stimulus. This ensures a one-to-one relation associating each stimulus with a precise emotion. Finally, the battery proposes a ranking of the stimuli among each emotion set (from the most obvious to the most ambiguous), which allows for choosing the emotional clarity of each stimulus included in a study, according to the desired difficulty of the task. To summarize, this battery is large (with various emotions, stimulus types and actors), highly validated (on the basis of an expert analysis followed by a large validation), highly standardized (controlled duration and intensity of the stimuli) and based on a qualitative evaluation (allowing a precise distinction between stimuli concerning the emotional intensity and specificity).

The potential applications of this battery are large and various. In the field of neuropsychology, it could be a useful tool to explore (by means of behavioural but also neuroimagery approaches) the cerebral correlates of emotional prosody, and for example (1) the dissociation between identity and emotion processing in the voice (e.g. Bedart & Belin, 2004), or (2) the dissociation between the processing of different emotions (e.g. Harciarek et al., 2006). Moreover, this battery, furnishing an auditory matching piece to the existing EFE batteries, could be a helpful tool to investigate the neuropsychological dissociations between visual and auditory processing of emotions, and particularly to explore the crossmodal processing of emotions (e.g. Grossman et al., 2006). Finally, as the exploration of emotional processing is a flourishing field in psychiatry, this battery could lead to several applications among clinical populations (e.g. Maurage et al., in press). These examples are of course not exhaustive, as this battery is designed to become a multi-purpose tool, potentially useful in a wide range of studies.

Nevertheless, this study is a first step in developing a more complete and global battery of emotional vocal expressions, and further works should develop this project in at least three directions. First, while the meaningless stimuli (i.e. interjections) are already suitable for studies in other languages, their validity among non French-speaking populations should be ascertained. Second, it should be noted that our semantic stimuli (i.e. words) can be used as emotional non-words among non-French speaking subjects, as the emotional features of the word are contained in the prosody and not in the word "paper" (which has a neutral meaning). Nevertheless, new single-word stimuli should be recorded in different countries, in order to obtain a comparable material in different languages. Finally, new stimuli, notably expressing other emotions, could be recorded and validated to expand the battery. We thus appeal to researchers studying emotional prosody in other languages, in order to develop this battery and we propose to include, in the material proposed on our website, any material validated in other languages, constructed according to the same procedure as presented in this paper.

To sum up, it appears crucial to develop a global, validated and freely available pool of emotional vocal stimuli, as the interest for the exploration of auditory processing of emotions increased exponentially during the last years, in normal as well as pathological populations. This study intend to be the first step towards this direction, by proposing a validated and standardized battery of emotional vocal stimuli (word and interjection), which could facilitate future studies using emotional vocal expressions, after the fashion of what has been conducted in the domain of the emotional facial expressions.

Acknowledgments: Frédéric Joassin is a Research Associate at the National Fund for Scientific Research (F.N.R.S., Belgium), and Pierre Maurage is a Research Assistant at the FNRS. The authors would like to Marie-Laure Martin and Aurélie Mullier for their support.

References

Ackerman, J.M., Shapiro, J.R., Neuberg, S.L., Kenrick, D.T., Becker, D.V., Griskevicius, V., Maner, J.K., & Schaller, M. (2006). They all look the same to me (unless they're angry): From out-group homogeneity to out-group heterogeneity. *Psychological Science*, 17, 836-840.

Banse, R., & Scherer, K.R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70, 614-636.

Bedard, C., & Belin, P. (2004). A "voice inversion effect?". *Brain and Cognition*, *55*, 247-249.

Blair, R.J., Morris, J.S., Frith, C.D., Perrett, D.I., & Dolan, R.J. (1999). Dissociable neural responses to facial expressions of sadness and anger. *Brain*, 122, 883-893.

Borod, J.C., Welkowitz, J., Alpert, M., Brozgold, A.Z., Martin, C., Peselow, E., & Diller, L. (1990). Parameters of emotional processing in neuropsychiatric disorders: Conceptual issues and a battery of tests. *Journal of Communication Disorders*, 23, 247-271.

Calvert, G.A. (2001) Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex*, 11, 1110-1123.

Campanella, S., Gaspard, C., Debatisse, D., Bruyer, R., Crommelinck, M., & Guérit, J.M. (2002). Discrimination of emotional facial expressions in a visual oddball task: An ERP study. *Biological Psychology*, 59, 171-186.

Davidson, R.J. (2003). Affective neuroscience and psychophysiology: Toward a synthesis. *Psychophysiology*, 40, 655-665.

Ekman, P., & Friesen, W.V. (eds.) (1976). *Pictures of Facial Affect*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., Sorenson, E.R., & Friesen, W.V. (1969). Pan-cultural elements in facial displays of emotions. *Science*, 164, 86-88.

Ethofer, T., Pourtois, G., & Wildgruber, D. (2006). Investigating audiovisual integration of emotional signals in the human brain. *Progress in Brain Research*, 156, 345-361.

Everhart, D.E., Demaree, H.A., & Shipley, A.J. (2006). Perception of emotional prosody: moving toward a model that incorporates sex-related differences. *Behavioural Cognitive Neuroscience Review*, 5, 92-102.

Fecteau, S., Armony, J.L., Joanette, Y., & Belin, P. (2005). Judgment of emotional nonlinguistic vocalizations: Age-related differences. *Applied Neuropsychology*, 12, 40-48.

Frijda, N.H., Kuiper, P., & Terschure, L. (1989). The relationships between emotion appraisal and emotion action readiness. *Journal of Personality and Social Psychology*, 57, 212-220.

Golan, O., Baron-Cohen, S., & Hill, J. (2006). The Cambridge Mindreading (CAM) Face-Voice Battery: Testing complex emotion recognition in adults with and without Asperger syndrome. *Journal of Autism and Developmental Disorders*, 36, 169-183.

Golan, O., Baron-Cohen, S., Hill, J.J., & Rutherford, M.D. (2007). The "reading the mind in the voice" test-revised: A study of complex emotion recognition in adults with and without autism spectrum conditions. *Journal of Autism and Developmental Disorders*, 37, 1096-1106.

Greimel, E., Macht, M., Krumhuber, E., & Ellgring, H. (2006). Facial and affective reactions to tastes and their modulation by sadness and joy. *Physiology and Behavior*, 89, 261-269.

Grossmann, T., Striano, T., & Friederici, A.D. (2006). Crossmodal integration of emotional information from face and voice in the infant brain. *Developmental Science*, 9, 309-315.

Harciarek, M., Heilman, K.M., & Jodzio, K. (2006). Defective comprehension of emotional faces and prosody as a result of right hemisphere stroke: modality versus emotion-type specificity. *Journal of International Neuropsychology Society*, 12, 774-781.

Hermans, D., Baeyens, F., Lamote, S., Spruyt, A., & Eelen, P. (2005). Affective priming as an indirect measure of food preferences acquired through odor conditioning. *Experimental Psychology*, 52, 180-186.

Hess, U., & Blairy, S. (1995). *Set of Emotional Facial Stimuli*. University of Quebec at Montréal, Department of Psychology, Montreal, Canada.

Hess, U., Kappas, A., & Scherer, K.R. (1989). Multi-channel communication of emotion: Synthetic signal production. In: Scherer, K.R. (ed.), *Facets of Emotion*. Hillsdale, NJ: Lawrence Erlbaum Associate, pp. 161-182.

Izard, C.E. (1977). *Human Emotions*. New York: Plenum.

Kappas, A., Hess, U., & Scherer, K.R. (1991). Voice and emotion. In: Feldman, R.S., Rimé, B. (eds.), *Fundamental of Non-verbal Behaviour*. New York: Cambridge University Press, pp. 200-238.

Kotz, S.A., Meyer, M., & Paulmann, S. (2006). Lateralization of emotional prosody in the brain: An overview and synopsis on the impact of study design. *Progress in Brain Research*, 156, 285-294.

Leventhal, H. (1984). A perceptual-motor theory of emotion. In: Berkowitz, L. (ed.), *Advances in Experimental Social Psychology*. New York: Academic Press, pp. 117-182.

Matsumoto, D., & Ekman, P. (1988). Japanese and Caucasian facial expressions of emotion and neutral faces (JACFEE and JACNeuF). Available from www.paulekman.com.

Maurage, P., Campanella, S., Philippot, P., Pham, T., & Joassin, F. (in press). The crossmodal facilitation effect is disrupted in alcoholism: A study with emotional stimuli. *Alcohol.*

Minzenberg, M.J., Poole, J.H., & Vinogradov, S. (2006). Social-emotion recognition in borderline personality disorder. *Comprehensive Psychiatry*, 47, 468-474.

Monnot, M., Nixon, S., Lovallo, W., & Ross, E. (2001). Altered emotional perception in alcoholics: Deficits in affective prosody comprehension. *Alcoholism-Clinical and Experimental Research*, 25, 362-369.

Montoya, P., & Sitges, C. (2006). Affective modulation of somatosensory-evoked potentials elicited by tactile stimulation. *Brain Research*, 1068, 205-212.

Oatley, K.J., & Johnson-Laird, P.N. (1987). Towards a cognitive theory of emotions. *Cognition and Emotion*, 1, 29-50.

Pell, M.D. (2002). Evaluation of nonverbal emotion in face and voice: Some preliminary findings on a new battery of tests. *Brain and Cognition*, 48, 499-504.

Pourtois, G., de Gelder, B., Bol, A., & Crommelinck, M. (2005). Perception of facial expressions and voices and of their combination in the human brain. *Cortex*, 41, 49-59.

Scherer, K.R. (1984). Emotion as a multicomponent process: A model and some cross-cultural data. In: Shaver, P. (ed.), *Review of Personality and Social Psychology.* Beverly Hills, CA: Sage, pp. 37-63.

Scherer, K.R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99, 143-165.

Electrocortical (EEG) correlates of music and states of consciousness

Lazar Skaric [1] - Milorad Tomasevic [2] - Dejan Rakovic [1, 3]
Emil Jovanov [6, 4] - Vlada Radivojevic [5] - Predrag Sukovic [1]
Marko Car [5] - Dejan Radenovic [1]

[1] *Faculty of Electrical Engineering, Belgrade, Yugoslavia*
[2] *Vinca Institute of Nuclear Sciences, Belgrade, Yugoslavia*
[3] *ECPD of the United Nations University for Peace, Belgrade, Yugoslavia*
[4] *Electrical and Computer Engineering Dept. The University of Alabama in Huntsville*
[5] *Institute for Mental Health, Department of Clinical Neurophysiology, Belgrade, Yugoslavia*
[6] *Institute for Biological Research, Center for Multidisciplinary Studies, Belgrade, Yugoslavia*

jovanov@ieee.org

Abstract

The study of the perception of music is a paramount example of multidisciplinary research. In spite of a lot of theoretical and experimental efforts to understand musical processing, attempts to localize musical abilities in particular brain regions were largely unsuccessful, save for the difference between musicians and non musicians, especially in hemispheric specialization and in EEG correlational dimensions. Having in mind that human emotional response to music and to art in general is limbic dependent, this motivated us to address our question to a similar possible neurobiological origin of musicogenic altered states of consciousness and its possible EEG correlates, "resonantly" induced by deep spiritual music. For example, as in sound-induced altered states of consciousness cultivated in some Eastern yogic practices. The musicogenic states of consciousness are evaluated within a group of 6 adults, upon the influence of 4 types of spiritual music. The most prominent changes in theta or alpha frequency bands were induced in two subjects, upon the influence of Indian spiritual music, Bhajan.

Keywords: Music Perception; States of Consciousness; EEG

Why do we like music? We all are reluctant, in music and art, to examine our sources of pleasure and strength. In part we fear success itself – that Understanding might spoil Enjoyment. And rightly so; Art often loses power when its psychological roots are exposed. No matter: when this happens we will go on, as always to seek more robust illusions! (Marvin Minsky)

1. Introduction

The study of the perception of music is a paramount example of multidisciplinary research, in which musicians, psychologists, neurobiologists, physicists, and engineers must communicate and work together. This study comprises three broad problem areas (Roederer, 1982): (a) perception of musical tones; (b) interpretation of acoustical information relevant to music; and (c) emotional response to musical messages. In the past two decades, a considerable mutual integration of these three problem areas has taken place, due to the progress in the understanding of general human brain functions, and the recognition that in the conscious state even the simplest perceptual events are bound to trigger operations that involve the brain as a whole.

The human brain does not appear to contain many new or drastically different processing centers when compared with the brains of any of our primate ancestors – save for cerebral hemisphere specialization, which is the basic evolutionary novelty of the human brain. This specialization is related to two quite different operational modes. One mode involves sequential analysis of single-channel information (such as required in language, speech, and thought processing, characteristic of the "dominant" hemisphere; the left hemisphere in about 97% of the subjects – the right-handed ones). The other involves a synthesis of many different parallel channels to accomplish the holistic determination of input stimuli (characteristic of the "minor" hemisphere).

Along these lines, it has been proposed that music is a language-like form by which humans express themselves and communicate with each other (Pribram, 1982). This analogy would then suggest that processing of musical indicants (such as melody and harmony) are predominantly "minor" hemisphere related, and that processing of musical symbols (such as hierarchically arranged phase structures) predominantly involves the "dominant" hemisphere. Semantic processing of both musical indicants and symbols should be related to the posterior cortical convexity. However, pragmatic processing of a user's musical experience and expression should be related to the frontolimbic cortical formations. And finally, syntactic processing of the arrangement of indicants and symbols should be related to the motor system of the brain, to which both posterior and frontal cortical formations project. However, it should be pointed out that the syntactic structure of music might be more dependent on semantic processing.

In spite of a lot of theoretical and experimental efforts to relate music and language processing (Pribram, 1982; Bernstein, 1976; Jackendoff & Lerdahl, 1982; Chomsky, 1980), attempts to localize musical abilities in particular brain regions were largely unsuccessful – both by electroencephalography (Petsche, Lindner, Rappelsberger & Gruber, 1988) and positron emission tomography (Sergent, 1993) – save for the difference between musicians (or musically talented subjects) and non-musicians:

(a) The processing of single musical notes and melodic line is represented in the "minor" hemisphere when presented to musically less sophisticated subjects, while it is equally well represented in both hemispheres of more sophisticated subjects.

(b) Less sophisticated subjects responded with a prop in brainwave complexity to rhythmical weakly chaotic music, while more sophisticated subjects showed higher EEG correlational dimensions (Birbaumer, Lutzenberger, Tau, Mayer-Kress & Braun, in press). It should also be added that neuroendocrinological measurements revealed a specific pattern of sexual hormones (increased testosterone in females; decreased in males) in composers and highly talented adolescents (Hassler & Birbaumer, 1988).

It should be also pointed out that one of the most profound consequences of the evolution of human brain functions (and human consciousness itself) has been the emergence of systematic postponing of behavioral goals and rearrangement of behavioral priorities. This led to conflicts between cortical functions and those of the limbic system. While the limbic system in animals is mostly activated by environmental and somatic input, in humans it can also respond to internally evoked images displayed on the cortex during the process of thinking. As motivation and emotion are integral manifestations of limbic function (assuring that all cortical processes are carried out so as to be of maximum benefit to the organism, through the extended reticular-thalamic activating system: Baars, 1988), in humans they can be triggered with no relationship to the current state of the environment. It is along this line that we should seek leads toward understanding the human emotional response to music and to art in general, when the messages therein seem to be of no obvious survival value (Roederer, 1982).

This motivated us to address our question to a similar possible neurobiological origin of musicogenic altered states of consciousness, induced by deep spiritual music of different cultures (Rouget, 1980), and its possible EEG correlates.

The analogous, more frequently used physical mechanism for sound-induced altered states of consciousness is an introspective repeating of a certain type of sound or "mantr", which is chosen so as to "resonate" with the structure of an individual's nervous system (Nader, 1995). The sound resonances within the human lobe would be then achieved through a formation of standing sound waves, with a principal harmonic (of ~ 1000 Hz) having its maximal amplitude in the centre of the lobe cavity, i.e. around the region of limbic system. This activity therefore induces the local stimulation of thalamic formation through some mechano-chemical receptors (to be still specified therein).

2. Method

2.1. Subjects

The study was carried out on 6 healthy adult volunteers. There was one male and five females, whose ages ranged from 18 to 29 years with a mean age of 25 years. All subjects were free of any medication. Prior to the experiment, the subjects were informed verbally about all aspects of the experimental procedure.

2.2. Music

Four types of spiritual music were provided to the subjects during the experiments: (1) Indian Bhajan in Sanskrit, (2) Byzantine Pasha Liturgy in Greek, (3) Maronite Song in Arabian, and (4) Mozart's Requiem in Latin.

2.3. Apparatus

Electroencephalographs were recorded in an electromagnetically shielded room by a MEDELEC 1A97 EEG machine, with lower and upper band-pass filter limits set at 0.5 Hz and 30 Hz, respectively. Ag/AgCl electrodes with impedance less than 5 kΩ were placed at 16 locations (F7, F8, T3, T4, T5, T6, Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2) according to the International 10-20 system with average reference. The EEG outputs were digitized with 12-bit precision at a sampling rate of 128 Hz per channel using A/D converter Data Translation 2801.

2.4. Procedure

The experiment was conducted in a sound-proof room, with only one music piece a day. Each recording session was divided into three sequential periods:

- relaxing 5 min with eyes closed;
- listening of the music 10 min;
- after listening, 5 min.

During those periods three samples, one minute each, were recorded for every subject. The EEG record was stored on a hard-disk.

2.5. Data analysis

The length of each EEG-trace was 60 s (7680-points). Time-varying EEG spectra (spectrograms) with 0.5 Hz resolution were calculated by the MATLAB program using a 256-point FFT algorithm performed on 2 s Hamming-windowed, half-overlapped epochs. An array of EEG partial power spectra for each subject and each derivation was computed by integration by the trapezoidal rule of the spectrogram over the three frequency bands: θ (from 4 to 8 Hz), α (from 8 to 13 Hz), and β_1 (13 to 18 Hz). The Wilcoxon matched pairs test and Mann-Whitney U-test were used to determine significant differences between the spectral arrays of the relaxing period and the spectral arrays of the meditation period. The coherence of spectral arrays was estimated using Welch's averaged periodogram method at 512-point (4 s) epochs of EEG data divided into 256-point (2 s) detrended Hamming-windowed subsets with 240-point overlap. Total coherence for the each frequency band was calculated using the same methods as those described by Levine et al. (Levine, Hebert, Haynes & Strobel 1977).

3. Results

In Table 1 the results of the Wilcoxon matched pairs test for medians of EEG power of all 16 channels, prior and during the listening of music are shown.

In most cases, during the listening of music, the EEG power decrease is observed in various frequency bands. In three cases (out of 20), a significant power increase in theta and alpha bands is registered, in accordance with an intense aesthetic experience in these cases; the two most prominent spectrograms and corresponding diagrams of the temporal changes of spectral power are shown in Figs. 1 and 2.

Table 1. The EEG power changes during the listening of music

	SUBJECT 1	SUBJECT 2	SUBJECT 3	SUBJECT 4	SUBJECT 5	SUBJECT 6
	band	band	band	band	band	band
music	θ α β_1	θ α β_1	θ α β_1	θ α β_1	θ α β_1	θ α β_1
1	- - 0	- - 0	- + 0	+ - 0	- - -	0 - -
2	0 - -	- - -	- - -	0 0 0	- - -	x x x
3	x x x	+ + -	- - -	0 - 0	0 - -	- - -
4	0 - -	0 - -	- - -	x x x	0 - 0	x x x

+ sign. increase; - sign. decrease; 0 sign. no changes; x sign. not recorded.

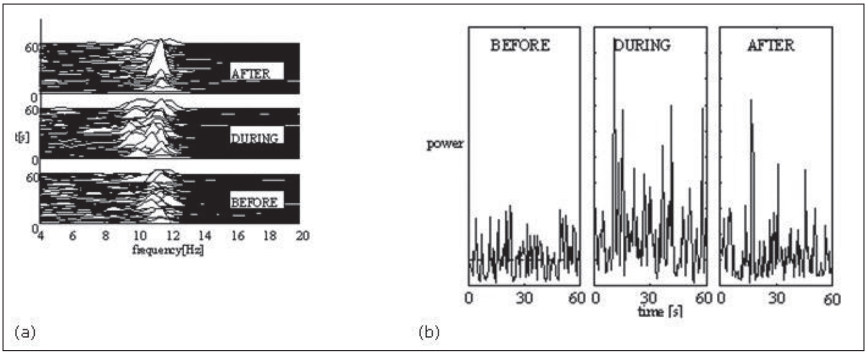


Figure 1. (a) The spectrogram with the observed EEG power increase in the alpha band and the appearance of slower alpha frequencies during the listening of music 1 in channel P3 of subject 3; (b) The corresponding temporal power changes in the alpha band

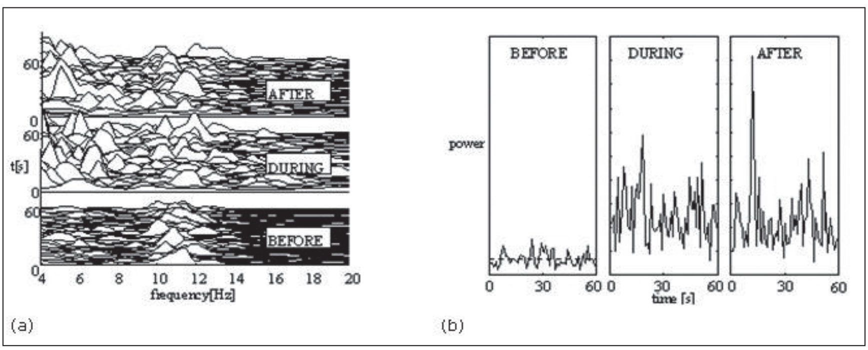


Figure 2. (a) The spectrogram with the observed high EEG power increas in the theta band during the listening of music 1 in channel T6 of subject 4; (b) The corresponding temporal power changes in the theta band

In Table 2 the results of the Mann-Whitney U-test for temporal arrays of the mean coherences of corresponding pairs of EEG channels are presented.

Table 2. The EEG coherence changes during the listening of music

		SUBJECT 1	SUBJECT 2	SUBJECT 3	SUBJECT 4	SUBJECT 5	SUBJECT 6
		band	band	band	band	band	band
music	cannels	$\theta\ \alpha\ \beta_1$	$\theta\ \alpha\ \beta_1$	$\theta\ \alpha\ \beta_1$	$\theta\ \alpha\ \beta_1$	$\theta\ \alpha\ \beta_1$	$\theta\ \alpha\ \beta_1$
1	F3-C3	0 0 -	0 0 0	0 0 0	+ - 0	0 0 0	0 - 0
	F4-C4	0 - 0	0 0 0	0 0 0	- - -	- - +	0 + 0
	F3-F4	0 0 0	0 0 0	0 0 0	0 - 0	- - +	0 - 0
	01-02	0 0 0	0 0 0	0 0 -	0 - 0	0 0 0	+ + 0
2	F3-C3	0 0 0	- 0 0	0 + 0	0 - 0	0 0 0	x x x
	F4-C4	0 0 0	0 0 0	0 + 0	0 0 0	0 0 0	x x x
	F3-F4	0 0 0	+ 0 0	0 + 0	0 0 +	0 0 0	x x x
	01-02	0 0 0	0 0 0	0 + 0	- 0 0	0 0 0	x x x
3	F3-C3	x x x	0 0 -	0 0 -	0 0 0	0 0 0	0 0 0
	F4-C4	x x x	0 0 -	0 0 0	0 - 0	0 0 +	0 + 0
	F3-F4	x x x	0 0 0	0 0 0	0 0 0	- 0 0	0 + 0
	01-02	x x x	0 0 0	0 + 0	0 0 0	+ 0 0	0 0 0
4	F3-C3	0 0 0	0 0 -	0 + 0	x x x	- - 0	x x x
	F4-C4	0 0 0	0 - 0	0 + 0	x x x	0 0 0	x x x
	F3-F4	0 0 0	0 0 0	0 0 0	x x x	- 0 +	x x x
	01-02	0 + 0	0 0 0	0 + 0	x x x	+ 0 0	x x x

+ sign. increase; - sign. decrease; 0 sign. no changes; x sign. not recorded.

It seems that changes in coherence during the listening of music are not correlated with aesthetic experiences. This might be a consequence of the observed increase in the mean ratio of EEG power of the right and the left hemispheres, the corresponding medians for all channels, and all subjects being R/L = 1.015 prior the music, and R/L = 1.082 during the listening of music. This fact is indicative, although the Wilcoxon matched pairs test did not give significant changes ($p = 0.09$).

4. Conclusion

According to our pilot study with six subjects and four types of spiritual music, it might be concluded that EEG power changes during their listening are quite individual. In the three cases where significant raise of power (i.e. relaxation) in theta and alpha bands is observed, the subjects have described their musical experiences as very pleasant – in contrast to the cases with drop in EEG power and unpleasant musical experiences. The most prominent changes were observed in subjects 3 and 4 upon the influence of the music 1 (Bhajan, Indian spiritual music, sung in Sanskrit); and somewhat less in subject 2 upon the influence of the music 3 (the Maronite spiritual music, sung in Arabian). Concerning the coherence increase, it seems not to be correlated with the aesthetic experience as, for instance, all subjects described their experience of music 2 as “slightly unpleasant”, while their coherence was even increased.

In spite of the observed particular EEG changes upon some types of spiritual music, it might be that more conclusive results could be achieved only in the case of more careful choice of subjects, regarding their musical affinities and/or education.

References

Baars, B.J. (1988). *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.

Bernstein, L. (1976). *The Unanswered Question: Six Talks at Harvard*. Cambridge, MA: Harvard University Press.

Birbaumer, N., Lutzenberger, W., Tau, H., Mayer-Kress, G., & Braun, C. (1996). Perception of music and dimensional complexity of Brain activity. *International Journal of Bifurcation and Chaos*, 6, 267-278.

Chomsky, N. (1980). *Rules and Representations.* New York: Columbia University Press.

Hassler, M., & Birbaumer, N. (1988). Handedness, music abilities, and dichaptic and dicjotic performance in adolescents: A longitudinal study. *Development Neuropsychology*, 4, 129-145.

Jackendoff, R., & Lerdahl, F. (1982). Grammatical parallel between music and language. In: Clynes, M. (ed.), *Music, Mind, and Brain: The Neuropsychology of Music*. New York: Plenum.

Levine, H., Hebert, J.R., Haynes, C.T., & Strobel, U. (1977). EEG coherence during the Transcendental Meditation technique. In: Orme-Johnson, D.W., & Farrow, J.T. (eds.), *Scientific Research of the Transcendental Meditation Program - Collected Papers*, vol. 1. Seelisberg: MERU Press, pp. 187-207.

Nader, T. (1995). *Human Physiology: Expression of Veda and the Vedic Literature*. Vlodrop: Maharishi.

Petsche, H., Lindner, K., Rappelsberger, P., & Gruber, G. (1988). The EEG: An adequate method to concretice brain process elicited by music. *Music Perception*, 6, 133-160.

Pribram, K.H. (1982). Brain mechanism in music: Prolegomena for a theory of the meaning of meaning. In: Clynes, M. (ed.), *Music, Mind, and Brain: The Neuropsychology of Music*. New York: Plenum.

Roederer, J.G. (1982). Physical and neuropsychological foundations of music: The basic questions. In: Clynes, M. (ed.), *Music, Mind, and Brain: The Neuropsychology of Music*. New York: Plenum.

Rouget, G. (1980). *La Musique et la Transe*. Paris: Gallimard.

Sergent, S. (1993). Mapping the musician brain. *Human Brain Mapping*, 1, 20-39.

Neuropsychological

Trends

Submission Procedure

Neuropsychological Trends (abbr. *Neuropsychol. trends*) is an electronic journal. Submissions are solicited twice a year (January and July). Notifications of acceptance will be published on the journal website approximately 3 months after each deadline, and all accepted papers will remain permanently published on the website.

All papers have to be submitted by e-mail to: neuropsychologicaltrends@psychoneuronet.com

The following items have to be attached as separate files:
- text file (included title page and abstract),
- tables (if any),
- figures (if any).

Papers shall be submitted in their final version.

Accepted paper formats include simple text, Microsoft Word, Rich Text Format (RTF). The name and affiliation of the author(s) and an electronic (e-mail) return address have to be indicated in all submissions.

Post-acceptance

PDF page proofs are sent to authors via e-mail only for the final check. No substantial modifications are allowed at this step.

With the submission to *Neuropsychological Trends* the authors acknowledge that reproduction of their paper will be allowed in case their paper is published.

Copyright