EDITORIAL

THE HIDDEN STRUCTURE OF NEUROPSYCHOLOGY: TEXT MINING OF THE JOURNAL *CORTEX*: 1991-2001

Ronald N. Kostoff[1], Henry A. Buchtel[2, 3], John Andrews[1] and Kirstin M. Pfeil[1]

([1]Office of Naval Research; [2]VA Ann Arbor Healthcare System; [3]University of Michigan Departments of Psychiatry and Psychology)

ABSTRACT

Background: The stated mission of *Cortex* is "the study of the inter-relations of the nervous system and behavior, particularly as these are reflected in the effects of brain lesions on cognitive functions." The purpose of this paper is to explore the relationship between the stated mission and the executed mission as reflected by the characteristics of papers published in *Cortex*. In addition, we examine whether the results and conclusions of an analysis of this kind are affected by the level of description of the published papers.

Objectives:
A) Identify characteristics of contributors to *Cortex*;
B) Identify characteristics of those who cite *Cortex*;
C) Identify recurring themes;
D) Identify the relationships among the recurring themes;
E) Compare recurring themes and determine their relationships to the mission of *Cortex*;
F) Identify the sensitivity of these results to the level of description of the *Cortex* papers used as the source database.
G) Compare *Cortex* characteristics with those of *Neuropsychologia*, another Europe-based international neuropsychology journal.

Methods: Text mining (extraction of useful information from text) was used to generate the characteristics of the journal *Cortex*. Bibliometrics provided the *Cortex* contributor infrastructure (author/ organization/ country/ citation distributions), and computational linguistics identified the recurring technical themes and their inter-relationships. Citation mining (the integration of citation bibliometrics and text mining) was used to profile the research user community. Four levels of published article description were compared for the analysis: Full Text, Abstract, Title, Keywords.

Results and Conclusions: Highly cited documents were compared among *Cortex*, *Neuropsychologia*, and *Brain*, and a number of interesting parametric trends were observed. The characteristics of the papers that cite *Cortex* papers were examined, and some interesting insights were generated. Finally, the document clustering taxonomy showed that papers in *Cortex* can be reasonably divided into four categories (papers in each category in parenthesis): Semantic Memory (151); Handedness (145); Amnesia (119); and Neglect (66).

It is concluded that *Cortex* needs to take steps to attract a more diverse group of contributors outside its continental Western European base if it wishes to capture a greater share of seminal neuropsychology papers. Further investigation of the critical citation differences reported in the paper is recommended.

Key words: information technology, text mining, bibliometrics, computational linguistics, citation mining, document clustering, neuropsychology

INTRODUCTION

The stated mission of *Cortex* is "the study of the inter-relations of the nervous system and behavior, particularly as these are reflected in the effects of brain lesions on cognitive functions." (See Journal Title Page) The aim of this paper is to examine the relationship between the stated mission and the executed mission as reflected by the characteristics of papers published in the journal. This was done by determining the technical and thematic characteristics of papers published in *Cortex*, and their inter-relationships as expressed by the categories in different taxonomies. In addition, we set out to ascertain the infrastructure (authors, institutions, countries) underlying the papers published in *Cortex*, as well as the infrastructure and technical focus of the community of authors who cite papers published in *Cortex*. Finally, we were interested in determining whether the results and conclusions about the technical themes and their relationships differ according to the level of information contained in the specific record field analyzed (Keywords, Titles, Abstracts or Full Text).

For the past decade, the first author has been developing ways to obtain the above types of information from large bodies of unstructured or semi-structured text (Kostoff, 1993, 1994, 1997a; Kostoff et al., 1995, 1998, 1999, 2000a, 2000b). These processes are collectively known as text mining (Hearst, 1999; Kostoff, 2003; Kostoff and DeMarco, 2001; Kostoff and Geisler, 1999; Losiewicz et al., 2000; Zhu and Porter, 2002). They consist of three generic components: information retrieval, information processing, and information integration. It was decided to apply text mining to obtain the different perspectives on *Cortex* outlined above. An iterative query development process is usually used for this kind of task (Kostoff et al., 1997), but this is not needed when analyzing a database of papers published in a particular journal. The main focus of the study was the application of information processing and information integration to a subset of the papers published in *Cortex*.

Methods

There are four components of the specific approach selected: database selection, bibliometrics analysis, citation mining and computational linguistics analysis. Each will be outlined.

I. Database Selection

Two databases were used for the study. The first database was the Web version of the Science Citation Index (SCI) (ISI Web of Science. Thomson ISI. Philadelphia PA USA), which consisted of all *Cortex* records from 1991-mid-2001 classified in the SCI as articles. Four hundred ninety-four records were retrieved, of which 481 were full articles with abstracts. Most of the records included authors, titles, author addresses, author keywords, abstract narratives, and references cited. An update of this database was used to compute prolific countries and references cited for articles published in *Cortex* (and *Neuropsychologia*) for 2002-mid-2004. Additionally, for the citation mining part of the study, the Web version of the SCI was used to extract all papers from 1994-2004 that cited *Cortex* papers published in 1993-1994.

The second database consisted of all of the 203 full text *Cortex* articles published from 1997 to 2000. These articles were supplied by the publisher in electronic format.

II. Bibliometric Analysis

The purpose of the bibliometrics analysis is to quantify the basic technical infrastructure of *Cortex*. This quantification is obtained through counting items such as authors, institutions, countries, and citations. While the quantification procedure is straight-forward, its interpretation can be quite complex.

The bibliometrics section has two components: Publication Bibliometrics (e.g., prolific authors and numbers of papers published); and Citation Bibliometrics (e.g., frequency of most cited documents). These are compared with similar results from the journal *Neuropsychologia*, and in one case, results from the journal *Brain* are included as well.

III. Citation Mining

Citation mining integrates citation bibliometrics and computational linguistics. Its purposes are to profile the documented user community, and to show the technical disciplines into which the cited research areas are evolving. In citation mining, a sample of papers describing the research area is selected, and all papers in the SCI that cite the sample papers are retrieved. Bibliometrics and computational linguistics are performed on this sample. The bibliometrics displays characteristics of the citing community, and the computational linguistics portrays the technical thrusts (and interrelationships) of the citing disciplines.

The sample selected consists of all articles published in *Cortex* in 1993-1994. There were a total of 73 papers selected. Over 1800 separate citing articles were retrieved. These 1800 citing articles were citation mined.

IV. Computational Linguistics Analysis

The purpose of the computational linguistics is to use the quantification of text patterns to identify the technical themes of the database, the relationship among those themes, and the relationship between the themes and the technical infrastructure revealed by this bibliometric analysis. The approach used in the present study was to count phrase combinations that co-occurred within bounded domains (e.g., Abstracts, specified numerical windows), and group documents that appeared in thematic clusters.

A. Taxonomy Generation: Statistical Clustering

General Analytic Approach

For the long-term *Cortex* analysis, the taxonomy of the Abstract field database covering *Cortex* papers from 1991-2001 was generated. Past text mining studies have used a variety of approaches to identify the main technical themes in the database. These include extracting key phrases and manually assigning them to categories; extracting key phrases and assigning them with statistical computer algorithm, using factor analyses and multi-link clustering; and grouping documents based on text similarity.

While factor analysis, multi-link phrase clustering, and document clustering were used for the present study, only document clustering will be reported in this paper. The complete detailed results and descriptions of methodologies can be found in Kostoff et al. (2004). The three techniques provided complementary perspectives on the structure of the *Cortex* literature. For the total SCI database, document clustering was performed using the Abstracts text only. In document clustering, documents are combined into groups based on their text similarity. Document clustering yields number of documents in each cluster directly, a proxy metric for level of emphasis in each taxonomy category.

Different document clustering approaches exist (Cutting et al., 1992; Guha et al., 1998; Hearst, 2000; Karypis et al., 1999; Karypis, 2004; Prechelt et al., 2002; Rasmussen et al., 1992; Steinbach et al., 2000; Willet, 1988; Wise, 1992; Zamir and Etzioni, 1998; Zhao and Karypis, in press). The approach presented in this section is based on a partitional clustering algorithm (Karypis [Web Document 2004]; Zhao, in press) contained within a software package named CLUTO. Most of CLUTO's clustering algorithms treat the clustering problem as an optimization process that seeks to maximize or minimize a particular clustering criterion function defined either globally or locally over the entire clustering solution space. CLUTO uses a randomized incremental optimization algorithm that is greedy in nature, and has low computational requirements.

CLUTO requires specification of the number of clusters desired. Cluster runs (of the total SCI database) of 32 clusters were generated. CLUTO also agglommorated the 32 clusters into a hierarchical tree (taxonomy) structure, and this taxonomy is presented in the clustering sections.

TABLE I
Most Prolific Authors – 1991-2002

Cortex		Neuropsychologia	
Author	Frequency	Author	Frequency
Mayes AR	11	Hodges JR	19
Carlesimo GA	10	Cowey A	16
Heilman KM	10	Grafman J	16
Pillon B	10	Heilman KM	15
Caltagirone C	9	Milner B	15
Dubois B	9	Rugg MD	14
Agid Y	8	Warrington EK	13
Denes G	8	Robbins TW	12
De Renzi E	8	Robertson IH	12
Grafman J	8	Bradshaw JL	11
Warrington -EK	8	Corballis MC	11
Capitani E	7	Frith CD	11
Sirigu A	7	Gazzaniga MS	11
Annett M	6	Parkin AJ	11
Basso A	6	Bryden MP	10
Cipolotti L	6	Driver J	10
Pizzamiglio L	6	Mayes AR	10
Sabbadini M	6	Patterson K	10
Umiltà C	6	Dolan RJ	9
Adair JC	5	Farah MJ	9

Results

I. Publication Bibliometrics

The first group of metrics consists of counts of papers published by different entities (e.g., authors, countries in which the work was carried out). These metrics can be viewed as output and productivity measures. They are not direct measures of research quality, although there is some threshold quality level inferred due to these papers' publication in the (typically) high caliber of journals accessed by the SCI.

A. Prolific Authors

Table I lists the twenty most prolific authors from *Cortex* and *Neuropsychologia* in this sample.

Of the twenty most prolific authors in *Cortex*, nine are from Italy, four are from the UK, four are from France, and three are from the USA. For *Neuropsychologia*, twelve are from the UK, four are from the USA, two are from Canada, one is from Australia, and one is from New Zealand. There are four names in common between the two lists (Mayes, Warrington, Heilman, Grafman). The first two authors are from the UK; the latter two are from the USA. The country distributions of the top twenty most prolific authors are very different, and different from those of other recent text mining studies performed by the first author. Almost half of the top performers in *Cortex* are from Italy, and Western Europe generally. The *Neuropsychologia* top performers are centered in the UK primarily, and in the countries of the TTCP (The Technical Cooperation Program) totally.

B. Prolific Organizations

Table II lists the twenty most prolific institutions. It should be noted that many different organizational components may be included under a single organizational heading (e.g., The University of Milan could include the Neurology Department, Neuropsychology Department, Neuroscience Department, etc.).

A number of observations from Table II follows. First, slightly under half of the twenty most prolific *Cortex* institutions are in academic settings (45%); the others are hospitals (25%) and research institutions (30%). It should be pointed out, however, that clinicians and researchers working in hospitals such as Hôpital La Pitie Salpêtriére (Paris) and The National Hospital for Neurology and Neurosurgery (Queen Square, London UK) usually have academic affiliations (London University, in the case of the National Hospital), so the institutional name may not identify the clinical versus academic status of the authors completely.

Second, a majority of the twenty most prolific *Neuropsychologia* institutions are in academic settings (60%); the others are hospitals (10%) and research institutions (30%).

Finally, seven of the top ten in the *Cortex* column are from continental Western Europe, namely, Italy and France. Contrast this with the *Neuropsychologia* column, where nine of the top ten are from the predominantly English speaking countries of USA, UK, and Canada.

TABLE II
Most Prolific Institutions – 1991-2001

Cortex		Neuropsychologia	
Institution	Frequency	Institution	Frequency
INSERM	38	MRC	65
IRCCS	30	Univ. Oxford	53
Univ. Padua	23	INSERM	51
Univ. Milan	20	Univ. London	44
Hop. La Pitie Salpetriere	17	Univ. Coll. London	42
Univ. Modena	17	Univ. Cambridge	40
CNRS	14	McGill Univ.	35
Natl. Hosp. Neurol. and Neurosurg.	13	Univ. Montreal	34
MRC	12	Inst. Neurol.	34
Univ. Florida	11	Inst. Psychiat.	31
Inst. Psychiat.	11	IRCCS	29
IRCCS S. Lucia	10	CNRS	27
Vet. Adm. Med. Ctr.	10	Harvard Univ.	26
Vet. Affairs. Med. Ctr.	10	Univ. Calif. Davis	25
Boston Univ.	9	Univ. St. Andrews	25
Hop. Henri Mondor	9	Univ. Milan	24
Univ. Coll. London	8	Univ. Calif. Los Angeles	23
Univ. Calif. Los Angeles	8	Natl. Hosp. Neurol. and Neurosurg.	23
Univ. Roma Tor Vergata	8	Radcliffe Infirm.	22
Univ. Oxford	8	Univ. Toronto	21

C. Prolific Countries

Table III lists the eighteen most prolific countries for the two time periods, 1991-2001, and the update 2002-2004. For 1991-2001, the USA has 35% greater relative representation in *Neuropsychologia* compared to *Cortex*, Italy is represented more than twice as often in *Cortex* compared to *Neuropsychologia*, whereas UK is represented about the same in both journals. However, as will be shown in the later analysis of most and least cited papers published in these journals, the prolific author/country results do not track the most cited paper results as well. In particular, the most cited papers published in *Cortex*, in the 1998-1999 sample examined, come from continental Western Europe, mainly Italy and France, and the USA and UK are not represented. The most cited papers published in *Neuropsychologia* come from the English-speaking countries, mainly UK and USA, reflecting the most prolific authors/ countries.

In both journals, the dominance of a handful of countries is clearly evident. In *Cortex*, three countries, USA, Italy, and UK are represented in 62% of the author address listings, while in *Neuropsychologia*, two countries, USA and UK, are represented in 53% of the author address listings. For the 2002-2004 update, the top layer in each journal is still dominated by the same countries, although Italy's share seems to have dropped somewhat in *Cortex*.

II. Citation Bibliometrics

The second group of metrics presented is counts of citations to papers published by different entities. While citations are often used as impact or quality

TABLE IIIa
Most Prolific Countries – 1991-2001

Cortex			Neuropsychologia		
Country	Frequency	Fraction	Country	Frequency	Fraction
UK	125	0.220	USA	427	0.273
Italy	117	0.206	UK	401	0.256
USA	113	0.199	Italy	137	0.087
France	57	0.100	Canada	133	0.085
Germany	32	0.056	Germany	121	0.077
Canada	30	0.053	France	90	0.057
Japan	17	0.030	Australia	49	0.031
Australia	15	0.026	Switzerland	30	0.019
Belgium	13	0.023	Israel	26	0.017
Netherlands	9	0.016	New Zealand	24	0.015
Austria	7	0.012	Netherlands	22	0.014
Israel	7	0.012	Japan	21	0.013
Finland	6	0.011	Belgium	19	0.012
Sweden	6	0.011	Spain	17	0.011
Spain	5	0.009	Sweden	14	0.009
Switzerland	4	0.007	Finland	14	0.009
Brazil	3	0.005	Denmark	12	0.008
Greece	2	0.004	Norway	9	0.006

TABLE IIIb
Most Prolific Countries – 2002-mid-2004

Cortex			Neuropsychologia		
Country	Frequency	Fraction	Country	Frequency	Fraction
UK	34	0.241	UK	166	0.258
USA	23	0.163	USA	161	0.250
Italy	20	0.142	Germany	56	0.087
Canada	8	0.057	Italy	53	0.082
Japan	8	0.057	France	49	0.076
France	7	0.050	Canada	47	0.073
Germany	7	0.050	Netherlands	25	0.039
Netherlands	7	0.050	Australia	20	0.031
Switzerland	5	0.035	Belgium	11	0.017
Australia	4	0.028	New Zealand	11	0.017
Belgium	4	0.028	Austria	9	0.014
Israel	4	0.028	Spain	9	0.014
Sweden	3	0.021	Switzerland	8	0.012
Austria	2	0.014	Japan	6	0.009
New Zealand	2	0.014	Finland	3	0.005
South Korea	2	0.014	Ireland	3	0.005
Finland	1	0.007	Norway	3	0.005
Norway	1	0.007	Sweden	3	0.005

metrics (Garfield, 1985), much caution needs to be exercised in their frequency count interpretation, since there are numerous reasons why authors cite or do not cite particular papers (Kostoff, 1998; Garfield, 1985; MacRoberts and MacRoberts, 1996).

The references (citations) in all the 494 retrieved papers were aggregated. Each paper citation was divided into author, year, and journal fields, and those cited most frequently were identified. The data were accumulated and presented in order of decreasing frequency. A small percentage of any of these categories received large numbers of citations.

A. Most Cited First Authors

Table IV lists the twenty most cited first authors.

There are thirteen authors in common between the two columns. Apart from Annett (*Cortex*) and Posner (*Neuropsychologia*), the differences are seen in the lower portion of the columns. These two authors reflect the areas of emphasis of the two journals, with handedness being the most most frequent category of articles in *Cortex* and cognition and attention being a frequent subject in *Neuropsychologia*. There is modest overlap between the most prolific authors and most cited first authors in both journals, being 20% in *Cortex* and 25% in *Neuropsychologia*. This modest overlap has been found in most text mining studies performed by the first author. It may be due to the time lag between an author's seminal works and present activity, the difficulty in being prolific and producing seminal papers, or prolific authors not being listed as first authors in papers that are co-authored by students and junior colleagues.

TABLE IV
Most Cited First Authors – 1991-2001 Database

Cortex		Neuropsychologia	
Author	Frequency	Author	Frequency
De Renzi E	286	Warrington EK	411
Warrington EK	250	Milner B	396
Annett M	158	Posner MI	386
Shallice T	148	Shallice T	284
Benton AL	107	Farah MJ	256
Milner B	100	De Renzi E	250
Bisiach E	94	Heilman KM	241
Wechsler D	90	Squire LR	229
Tulving E	86	Wechsler D	225
Squire LR	82	Bisiach E	218
Heilman KM	76	Kinsbourne M	216
Kinsbourne M	76	Schacter DL	182
Farah MJ	74	Petrides M	181
Schacter DL	74	Tulving E	176
Geschwind N	73	Sergent J	175
Kopelman MD	73	Kimura D	174
Kapur N	69	Kosslyn SM	171
Bryden MP	69	Benton AL	170
Nelson HE	67	Bradshaw JL	170
Caramazza A	65	Damasio AR	169

The latter cause may be significant. For example, the ten most recently published papers (all journals) of the five most prolific authors in *Neuropsychologia* (from Table I), and the five most prolific authors in Cortex (excluding Heilman, who is captured in the *Neuropsychologia* group) were examined. These ten authors in Table I were first authors in only 9 of the 100 total papers, similar to the experience for most of the other text mining studies conducted by the first author. For those prolific authors who are rarely first authors, the chances that they will accumulate significant first author citations are low.

B. Most Cited Documents

The twenty most cited documents in *Cortex* and in *Neuropsychologia* are listed in Table V for 1991-2001 and for the 2002-mid-2004 update. There are eleven documents in common between the two lists for 1991-2001, and five documents in common between the two lists for 2002-mid-2004. The latter five are a subset of the former eleven. The top ten document citations center around 1980 for the 1991-2001 database. For *Cortex*, the most cited papers and books range from 1966 to 1991 (from 25-0 years before the start of the survey), with the median being 1980. For *Neuropsychologia*, the most cited papers and books range from 1964-1990, with the median being 1982. For the updated 2002-2004 database, the top ten average date of publication for *Cortex* is 1981, and the top ten average for *Neuropsychologia* is 1984. Even though the referencing papers for *Cortex* are almost a decade later for the 2002-2004 papers relative to the 1991-2001 papers, the average of the top ten referenced papers increases only from 1979 to 1981. For *Neuropsychologia*, the parallel increase in the average year of publication moves from 1981 to 1984. Additionally, the articles published in *Brain* from 2002-2004 were retrieved, and the corresponding top ten average was 1984. An outlier published by Hoehn in 1967 helped depress this top ten average reference age in *Brain*.

The top ten most cited references from the *Brain* database were analyzed further. The three most recent of the ten were compared to the three least recent. The three most recent, published in the mid to late 1990s, included two papers on the mapping and analysis of images (especially PET and MRI images), and one paper on the pathological analysis of brain tissue using advanced techniques. The three least recent, published in the mid-60s to mid-70s, included two papers describing questionnaires to measure handedness and cognitive states of patients, respectively, and one paper with clinical descriptions of Parkinson's disease based on patient observations. Thus, the most recent of the highly cited papers from the *Brain* sample tend to focus on the higher tech 'hard' technologies, while the least recent tend to focus on the 'soft' clinical behavioral and observational technologies.

Because of space limitations, only the ten most cited *Cortex* papers from 1991-2001 will be summarized briefly. The Folstein paper describes the popular MMSE (Mini-Mental State Exam). The Oldfield paper describes a paper-and-pencil questionnaire for measuring the degree of right- and left-handedness. The Warrington reference refers to the Warrington Recognition Memory Test for Faces, which uses fifty target faces, but includes a considerable amount of non-facial information in its stimuli. The Snodgrass paper provides stimuli for use in studies in which the complexity and other characteristics of the stimulus need to be known. The Shallice book describes the connection between cognitive psychology and neuropsychology. The McKhann paper reports the consensus statement defining the characteristics of possible and probable Alzheimer's Disease The Shallice and Evans paper describes the tendency of patients with frontal lobe lesions to make errors when asked to make estimates of sizes and frequencies that are not generally known with certainty (e.g., how many camels are there in Holland). The Nelson paper introduced a simplified version of the Wisconsin Card Sorting Test appropriate for elderly patients or patients likely to be fatigued by the original version. The Albert paper concerns the dissociation of neglect as a disorder from visual field defects. The De Renzi paper addresses normative data and the screening power of a shortened version of the Token Test. The Annett book addresses the demographics and putative genetics of handedness. Overall, the most cited Cortex reference documents tend to focus on clinical behavioral tests, as opposed to surgical experiments, invasive diagnostic experiments, animal laboratory tests, or even non-invasive experiments, to any significant degree.

Taken together, this list of most cited papers has no single theoretical or conceptual theme, but papers on classification, methods and tests are

TABLE Va

Cortex – Most Cited Papers – 1991-2002 Database

Document	Frequency
Folstein MF, 1975, *J Psychiat Research*, Vol. 12, P189	52
Oldfield RC, 1971, *Neuropsychologia*, Vol. 9, P97	41
Warrington EK, 1984, *Recognition Memory Test*	35
Snodgrass JG, 1980, J *Exp Psychol Human*, Vol. 6, P174	34
Shallice T, 1988, *From Neuropsychology to Mental Structure*	34
Mckhann G, 1984, *Neurology*, Vol. 34, P939	33
Shallice T, 1978, *Cortex*, Vol. 14, P294	31
Nelson HE, 1976, *Cortex*, Vol. 12, P313	30
Albert ML, 1973, *Neurology*, Vol. 23, P658	24
De Renzi E, 1978, *Cortex*, Vol. 14, P41	23
Annett M, 1985, *Left Right Hand Brain*	23
Wechsler D, 1981, *Wechsler Adult Intelligence Scale*	22
De Renzi E, 1987, *Cortex*, Vol. 23, P575	21
Warrington EK, 1984, *Brain*, Vol. 107, P829	18
Annett M, 1970, *Brit J Psychol*, Vol. 61, P303	18
Warrington EK, 1991, *Visual Object Space*	17
Benton AL, 1983, *Contributions to Neuropsychological Assessment*	17
Milner B, 1971, *Brit Medical B*, Vol. 27, P272	17
Bruce V, 1986, *Brit J Psychol*, Vol. 77, P305	16
Luria AR, 1966, *Higher Cortical Functions in Man*	16

TABLE Vb

Cortex – Most Cited Papers – 2002-mid-2004 Database

Document	Frequency
Folstein MF, 1975, *J Psychiat Research*, Vol. 12, P189	14
Oldfield RC, 1971, *Neuropsychologia*, Vol. 9, P97	12
Bisiach E, 1978, *Cortex*, Vol. 14, P129	10
Bisiach E, 1990, *Neurology*, Vol. 40, P1278	8
Heilman KM, 1979, *Ann Neurol*, Vol. 5, P166	7
Albert ML, 1973, *Neurology*, Vol. 23, P658	7
Stuss DT, 1986, Frontal Lobes 6 Riddoch MJ, 1983, *Neuropsychologia*, Vol. 21, P589	6
Snodgrass JG, 1980, *J Expt Psychol Human*, Vol. 6, P174	6
Rossetti Y, 1998, *Nature*, Vol. 395, P166	5
Vallar G, 1998, *Trends Cogn Sci*, Vol. 2, P87	5
Smith EE, 1999, *Science*, Vol. 283, P1657	5
Wechsler D, 1981, *Wechsler Adult Intelligence Scale*	5
Bisiach E, 1996, *Brain 3*, Vol. 119, P851	5
Shallice T, 1988, *From Neuropsychology to Mental Structure*	5
Bisiach E, 1998, *Brain Cognition*, Vol. 37, P369	5
Bisiach E, 1998, *Conscious Cogn*, Vol. 7, P327	5
Wechsler D, 1987, *Wechsler Memory Scale*	5
Duncan J, 1999, *J Exp Psychol Gen*, Vol. 128, P450	5
Vallar G, 1986, *Neuropsychologia*, Vol. 24, P609	5

TABLE Vc

Neuropsychologia – Most Cited Papers – 1991-2001 Database

Document	Frequency
Oldfield RC, 1971, *Neuropsychologia*, Vol. 9, P97	141
Folstein MF, 1975, *J Psychiat Res*, Vol. 12, P189	98
Shallice T, 1988, *From Neuropsychology to Mental Structure*	78
Snodgrass JG, 1980, *J Exp Psychol Human*, Vol. 6, P174	75
Wechsler D, 1981, *Wechsler Adult Intelligence Scale*	65
Posner MI, 1984, *J Neurosci*, Vol. 4, P1863	58
McKhann G, 1984, *Neurology*, Vol. 34, P939	55
Talairach J, 1988, *Coplanar Stereotaxic Atlas of the Human Brain*	55
Nelson HE, 1976, *Cortex*, Vol. 12, P313	54
Warrington EK, 1984, *Recognition Memory Test*	50
Posner MI, 1990, *Annu Rev Neurosci*, Vol. 13, P25	49
Mesulam MM, 1981, *Ann Neurol*, Vol. 10, P309	46
Posner MI, 1980, *Q J Expt Psychol*, Vol. 32, P3	44
Warrington EK, 1984, *Brain*, Vol. 107, P829	44
Milner B, 1971, *Brit Medical B*, Vol. 27, P272	43
Kucera H, 1967, *Computational Analysis of Present-Day American English*	40
Milner B, 1964, *Frontal Granular Cor*, P313	39
Albert ML, 1973, *Neurology*, Vol. 23, P658	39
Stuss DT, 1986, *Frontal Lobes*	38
Goldmanrakic PS, 1987, *Hdb Physl 1*, Vol. 5, P373	37

TABLE Vd

Neuropsychologia – Most Cited Papers – 2002-mid-2004 Database

Document	Frequency
Folstein MF, 1975, *J Psychiat Res*, Vol. 12, P189	51
Oldfield RC, 1971, *Neuropsychologia*, Vol. 9, P97	45
Talairach J, 1988, *Coplanar Stereotaxic Atlas of the Human Brain*	35
McKhann G, 1984, *Neurology*, Vol. 34, P939	23
Snodgrass JG, 1980, *J Exp Psychol Human*, Vol. 6, P174	22
Friston KJ, 1995, *Human Brain Mapping*, Vol. 2, P189	22
Wechsler D, 1981, *Wechsler Adult Intelligence Scale*	21
Alexander GE, 1986, *Annual Rev Neurosci*, Vol. 9, P357	21
Kanwisher N, 1997, *J Neurosci*, Vol. 17, P4302	21
Posner MI, 1984, *J Neurosci*, Vol. 4, P1863	20
Nelson HE, 1976, *Cortex*, Vol. 12, P313	19
Warrington EK, 1984, *Recognition Memory Test*	19
Hoehn MM, 1967, *Neurology*, Vol. 17, P427	18
Milner AD, 1995, *Visual Brain Action*	17
Shallice T, 1988, *From Neuropsychology to Mental Structure*	17
Wechsler D, 1987, *Wechsler Memory Scale*	17
Aggleton JP, 1999, *Behav Brain Sci*, Vol. 22, P425	15
Cabeza R, 2000, *J Cognitive Neurosci*, Vol. 12, P1	15
Ekman P, 1976, *Pictures of Facial Affect*	15
Ungerleider LG, 1982, *Analysis of Visual Behav*, P549	14

clearly represented more often than one would expect by chance. In particular, the paper by Folstein et al. (1975) is frequently cited because the test described in that paper (MMSE) is used across disciplines as an approximate measure of cognitive status. The test requires little training to administer and covers several domains relevant to evidence of cognitive decline (orientation to person and place, memory, comprehension, calculations and internal mental manipulations, and the like). The exceptions to the "methods" theme are the books by Shallice and by Annett and the paper by Shallice and Evans (1978). While a task is described in the Shallice and Evans paper, that particular version of the test is not in common use, but the concept of executive function described in this paper continues to be important in neuropsychology (a modified version of the method, with norms, was published later by Axelrod et al., 1994).

C. Most Cited Journals

Table VI lists the top twenty most cited journals and their frequencies.

The journals at the top of the lists are mainly neurology and neuropsychology journals, with the exception of the general science journals *Science* and *Nature*. The bottom of the lists includes psychology journals that publish papers relevant to neuropsychology. In agreement with the emphasis on clinical findings and their significance, there are no highly cited journals specializing in basic genetics, biology or biochemistry. The first entries on the most cited journals lists that reflect other fields of science are, for *Cortex*, *Journal of the Acoustics Society of America (JASA)* and *American Journal of Medical Genetics* (numbers 110 and 176 on the list of journals, respectively), and, for *Neuropsychologia*, *JASA*, *Biological Cybernetics*, and *Cell* (numbers 71, 230, and 459 on the list of journals, respectively).

There are sixteen journals in common between the two lists. Add to this the majority commonality in Cortex and *Neuropsychologia* of most cited authors and most cited documents shown previously, and it can be concluded that both journals draw heavily upon the same intellectual heritage. Given the many similarities in intellectual heritage, and the modest similarities in production demographics (prolific authors, institutions), how do the papers published in the two journals impact the larger technical community? This question is partially answered in the next section.

TABLE VI

Most Cited Journals – 1991-2001 Database

Cortex		Neuropsychologia	
Journal	Frequency	Journal	Frequency
Neuropsychologia	1592	*Neuropsychologia*	5293
Cortex	1342	*Brain*	2228
Brain	807	*Cortex*	1690
Brain Lang	572	*Neurology*	1204
Neurology	515	*Brain Lang*	1062
J Neurol Neurosur Ps	447	*Science*	916
Cognitive Neuropsych	406	*Nature*	898
Brain Cognition	403	*J Neurosci*	869
Arch Neurol-Chicago	353	*J Neurol Neurosur Ps*	855
J Clin Exp Neuropsyc	270	*Brain Cognition*	832
Science	221	*Arch Neurol-Chicago*	777
Nature	185	*Cognitive Neuropsych*	728
Ann Neurol	160	*Exp Brain Res*	684
J Exp Psychol Learn	158	*J Neurophysiol*	636
Brit J Psychol	155	*J Cognitive Neurosci*	605
J Neurosci	138	*J Clin Exp Neuropsyc*	535
Psychol Rev	117	*J Exp Psychol Learn*	522
Psychol Bull	115	*Psychol Rev*	478
Cognition	114	*J Exp Psychol Human*	469
Percept Motor Skill	109	*Ann Neurol*	459

D. Citation Comparison among Cortex, Neuropsychologia, *and* Brain

To further compare citations among papers published in the three top cited journals above, *Cortex*, *Neuropsychologia*, and *Brain*, the following experiment was run. All articles published in *Cortex*, *Neuropsychologia*, and *Brain* in the years 1998-1999 were retrieved from SCI. There were 110 *Cortex* articles, 278 *Neuropsychologia* articles, and 341 *Brain* articles. Then, the ten most cited articles from each retrieval (the citations from each paper used for the tabulation of most and least cited are those listed in the SCI Times Cited field, and are the total citations received by each paper from all other papers in the SCI) were extracted, as well as the ten least cited articles, and various characteristics compared. The results are shown in Table VII.

Code: Type
Behav = Clinical Behavior Studies
Surgery = Surgical Interventions
Diag-Ni = Non-Invasive Diagnostic Tests
Diag-Inv = Invasive Diagnostic Tests

A number of interesting observations may be made from Table VII. First, the most cited articles in *Neuropsychologia* are cited, on average, more than three times as often as the most cited articles in *Cortex*, and the most cited articles in *Brain* are cited, on average, more than twice as often as the most cited articles in *Neuropsychologia*.

Second, the most cited papers have more authors than the least cited, in all three journals, and the effect is most pronounced in *Neuropsychologia*. Additionally, the average number of authors increases with the average number of citations, ranging from about four authors of the most cited *Cortex* papers to about seven authors of the most cited *Brain* papers.

Third, the most cited papers have substantially more references than the least cited, in both journals, and the effect is most pronounced in *Neuropsychologia*. Additionally, the average number of citations increases with the average number of references (an effect observed by the first author in recent unpublished text mining studies), ranging from about 46 references in the most cited *Cortex* papers to about 68 references in the most cited *Brain* papers.

Fourth, there is no clear overall trend in citations as a function of institutional representation. The institution/ (institution + university) ratio (where institution in the table cells should be interpreted as any non-university organization; e.g., research laboratory, clinic, hospital, company) for most cited papers starts at 0.5 for Cortex, drops to 0.2 for *Neuropsychologia*, and increases sharply to 0.8 for *Brain*. This ratio for least cited papers starts at 0.4 for both *Cortex* and *Neuropsychologia*, and decreases to 0.2 for *Brain*. Its most dramatic change is from 0.8 for the most cited *Brain* papers to 0.2 for the least cited *Brain* papers.

Fifth, the most cited papers in *Cortex* are all from continental Western Europe, with heavy representation from Italy and France, while the least cited papers in *Cortex* represent four different continents. The most cited papers in

TABLE VII
Comparison of Most Cited/Least Cited Papers, Published 1998-1999

	Cortex		Neuropsychologia		Brain	
	Most Cited	Least Cited	Most Cited	Least Cited	Most Cited	Least Cited
N. Auth						
Average	3.9	2.8	5.2	2.6	7.1	4.6
Median	4	3	5	1	7.5	4.5
N. Refs						
Average	46.3	28	52.5	26.8	68.3	42.4
Median	49	29.5	49	26	62.5	35
N. Cites						
Average	21	0.8	71.3	0	166.8	2.8
Median	18.5	1	67.5	0	157	3
Org						
Institution	5	4	2	4	8	2
University	5	6	8	6	2	8
Country	4 Italy 3 France 1 Austria 1 Belgium 1 Germany 1 Australia	2 Italy 2 USA 2 Germany 2 Japan 1 Neth	4 UK 4 USA 1 Italy 1 Canada	5 USA 2 Italy 1 NZ 1 Neth 1 Australia	5 UK 2 USA 2 Canada 1 Germany	3 Japan 1 USA 1 UK 1 France 1 Italy 1 Canada 1 Germany 1 Neth
Type						
Behavior	8		4			
Surgery			1		2	
Diagnostic-NI	2		5		7	
Diagnostic-INV					1	

Neuropsychologia are, with the exception of Italy, from the UK and North America (with heavy representation from the UK and USA), while the least cited papers have more representation from Western Europe but none from the UK. The most cited papers in *Brain* are from the major English-speaking countries, whereas the least cited are scattered around Western Europe, Asia, and North America.

Sixth, there is a distinct shift in type of study (the bottom of Table VII) in proceeding from *Cortex* to *Neuropsychologia* to *Brain*. Clinical behavioral studies, many of them essentially case studies, predominate the most cited Cortex papers. There are only two papers characterized as Diagnostic-Non-Invasive (e.g., PET, MRI, etc). *Neuropsychologia* has more of a balance between Behavioral and Diagnostic-Non-Invasive in its ten most cited papers. *Brain* shows a heavy emphasis on Diagnostic-Non-Invasive (7/10), two papers on surgical procedures, and one on Diagnostic-Invasive. Based on reading Abstracts from each of these journals, the types as represented in the top ten most cited articles roughly approximate the types of papers published overall. Thus, as citations increase in absolute amounts, the study type transitions from the clinically oriented behavioral focus to the correlates with more objective measurements. Also, as the results from the most cited papers section showed, as the study type transitions from the clinically oriented behavioral focus ('soft' technology) to the more objective measurements ('hard' technology), the most cited papers tend to become more recent.

III. Citation Mining

A. Bibliometrics

For the 73 *Cortex* sample papers published in 1993-94, there were over 1800 citing papers. There were a total of 3424 citing authors, the top ten of which are shown in Table VIII.

As shown in Table IX, the 73 *Cortex* sample papers were cited in 322 different journals.

As shown in Table X, the citing authors came from 1046 institutions. The highest frequency institutions are split between Western Europe and the USA.

TABLE VIII

Authors of Cortex Citing Papers, 1993-1994 Database

Rank	N. Records	Author
1	30	Gabrieli JDE
2	26	Hodges JR
3	22	Markowitsch HJ
4	20	Della Sala S
5	17	Annett M
6	17	Capitani E
7	16	Heilman KM
8	16	Spinnler H
9	15	Caramazza A
10	15	Laiacona M

TABLE IX

Journals Publishing Papers that Cite Cortex, 1993-1994 Database

Affiliation (Journal)
Total Journals: 322
Top 10

	N. Records	Journal
1	179	*Neuropsychologia*
2	149	*Cortex*
3	80	*Neurocase*
4	77	*Neuropsychology*
5	72	*Cognitive Neuropsychology*
6	63	*Brain and Cognition*
7	63	*Brain and Language*
8	51	*Brain*
9	50	*Journal of Clinical and Experimental Neuropsychology*
10	28	*Journal of Neurology Neurosurgery and Psychology*

TABLE X

Organizations of Cortex Citing Authors, 1993-1994 Database

Affiliation (Organization)
Total Organizations: 1046
Top 10

	N. Records	Affiliation (Organization)
1	52	Univ. Milan
2	47	IRCCS
3	46	Boston Univ.
4	44	Vet. Affairs Med. Ctr.
5	43	MRC
6	39	Harvard Univ.
7	39	Univ. Toronto
8	36	Univ. Cambridge
9	34	Univ College London
10	32	Univ. Calif. San Diego

TABLE XI

Countries of Cortex Citing Authors, 1993-1994 Database

Affiliation (Country)
Total Countries: 45
Top 10

	N. Records	Affiliation (Country)
1	511	USA
2	343	UK
3	171	Italy
4	110	France
5	100	Canada
6	86	Germany
7	52	Australia
8	45	Japan
9	36	Belgium
10	34	Switzerland

As shown in Table XI, the highest frequency citing countries are the USA and Western Europe.

As shown in Table XII, 19613 different first authors were cited in the papers that cited the *Cortex* sample papers.

As shown in Table XIII, the *Cortex* sample citing papers cited a total of 46627 different books and papers. The *Cortex* papers in the top ten were by De Renzi and Kapur. The top ten citations are mainly to journal articles (9/10).

Finally, there were 8966 journals cited by the citing papers of the sample *Cortex* papers. The top ten are shown in Table XIV. Based on all the

TABLE XII

First Authors Cited by Cortex Citing Authors, 1993-1994 Database

Cited First Authors
Total Cited First Authors: 19613
Top 10

	N. Instances	Cited First Author
1	1254	Warrington, Elizabeth (UK)
2	855	De Renzi, Ennio (Italy)
3	828	Squire, Larry (USA)
4	822	Schacter, Daniel (USA)
5	699	Kapur, N (UK)
6	694	Markowitsch HJ (Germany)
7	688	Tulving, Endel (Canada)
8	658	Hodges, JR (UK)
9	644	Annett, M (UK)
10	628	Kopelman, MD (UK)

TABLE XIII

Documents Cited by Cortex Citing Authors, 1993-1994 Database

Cited Papers
Total Cited Books and Papers: 46627
Top 10

	N. Instances	Cited Paper
1	247	Folstein MF, 1975, *J Psychiat Res*, Vol. 12, P189
2	210	Warrington EK, 1984, *Brain*, Vol. 107, P829
3	189	Snodgrass JG, 1980, *J Expt Psychol Human*, Vol. 6, P174
4	161	De Renzi E, 1994, *Cortex*, Vol. 30, P3
5	156	Warrington EK, 1987, *Brain*, Vol. 110, P1273
6	154	Mckhann G, 1984, *Neurology*, Vol. 34, P939
7	147	Kapur N, 1995, *Cortex*, Vol. 31, P99
8	131	Oldfield RC, 1971, *Neuropsychologia*, Vol. 9, P97
9	130	Shallice T, 1988, *From Neuropsychology to Mental Structure*
10	129	Warrington EK, 1983, *Brain*, Vol. 106, P859

bibliometrics results, citation mining and non-citation mining, there appears to be a symbiotic relationship among *Cortex*, *Neuropsychologia*, and *Brain*. Of the journals in the list, these three are the most relevant to neuropsychology and are the journals most likely to be read by contributors to *Cortex*. Each, however, has its own niche or area of focus. Of the three journals, *Neuropsychologia* tends to be the journal for manuscripts on brain function in normal individuals. *Cortex*, as indicated by its mission statement, attracts papers about neuropsychological findings in patients with brain damage. *Brain* has a much wider range of articles, though it tends not to publish papers about brain function in normal individuals.

B. Computational Linguistics

A taxonomy based on the citing papers was generated, using factor analysis. Among the high frequency phrases, there are no new applications within the central discipline identified, or research and applications external to the central discipline identified. Of course, this is not surprising, given that most of the citing papers were themselves published in *Cortex*.

TABLE XIV

Journals Cited by Cortex Citing Authors, 1993-1994 Database

Cited Journals
Total Cited Journals: 8966
Top 10

	N. Instances	Cited Journal
1	7559	*Neuropsychologia*
2	6095	*Cortex*
3	3822	*Brain*
4	2993	*Cognitive Neuropsychology*
5	2944	*Neurology*
6	2223	*J Neurol Neurosurg Psychiatry*
7	2186	*Brain Language*
8	2047	*Arch. Neurol-Chicago*
9	1936	*Nature*
10	1816	*Brain Cognition*

IV. Computational Linguistics

Taxonomy Generation: Document Clustering

The 481 *Cortex* articles with Abstracts were clustered by the CLUTO algorithm into 32 elemental groups, yielding a high resolution average of fifteen records per group. These elemental groups were aggregated into different hierarchical levels. In the highest level, the 481 records were divided thematically into two categories; in the next highest level, the 481 records were divided into four categories; and so on. In the following analysis, the first three levels will be analyzed, and the themes (and associated numbers of records) of each category will be presented and discussed. The numbers in parentheses after the themes are the numbers of records.

CLUTO divides Level 1 into two categories: Handedness/Awareness (211) and Memory (270).

CLUTO divides Level 2 into four categories, by dividing each Level 1 category into two sub-categories. Handedness/Awareness (211) is divided into Handedness (145) and Neglect (66). Memory (270) is divided into Semantic Memory (151) and Amnesia (119).

CLUTO divides Level 3 into eight categories, by dividing each Level 2 category into two sub-categories. Handedness (145) is divided into Lateral Classification (82) and Lateral Movement (63). Neglect (66) is divided into Visual Field Neglect (38) and Neglect Diagnostics (28). Semantic Memory (151) is divided into Verbal/Numerical (76) and Visual/Spatial (75). Amnesia (119) is divided into Amnestic Symptoms (50) and Physiology of Amnesia (69).

The following *Cortex* flat taxonomy can be generated. The bullets under each category represent the 32 elemental cluster themes. The numbers in parentheses after the top level categories and before the themes are the numbers of records.

HANDEDNESS (145)
Lateral Classification (82)
(13) – Selective attention.

(12) – Ear asymmetry, especially in detecting/identifying dichotic stimuli.

(14) – Childhood dyslexia, especially associated with deficits in inter-hemispheric interactions, as well as visual and language deficits.

(12) – Immune and familial genetic disorders, emphasizing relation to laterality and handedness.

(16) – Handedness experiments, and relation of handedness to other variables.

(15) – Hand preferences, and the relationship of asymmetries to skills.

Lateral Movement (63)

(15) – Hand movements, especially manual asymmetries, for diagnosing apraxia.

(11) – Handedness, especially in relation to motor functions, such as turning direction, reaching, grasping, both intra- and inter-manual.

(10) – Threshold detection, especially for hearing sounds, with some associations to simultaneous stimuli and bimanual tasks.

(13) – Emotional stimuli, and hemispheric arousal related to facial expressions.

(14) – Hemispheric response differences to mainly visual stimuli, including color.

NEGLECT (66)

Visual Field Neglect (38)

(15) – Visual field stimuli, including dots and letters, emphasizing lateral imagery experiments.

(11) – Extinction, emphasizing tactile but including other sensory inputs, and neglect, using contra-lesional and ipsi-lesional data.

(12) ´ Neglect in brain damaged patients, emphasizing right brain damage,

Neglect Diagnostics (28)

(13) – Line bisection tests, for evaluating neglect.

(15) – Neglect, including personal and extra-personal, emphasizing cancellation experiments and left neglect.

SEMANTIC MEMORY (151)

Verbal/Numerical (76)

(16) – Arithmetic and numerical calculations and facts, including coding tasks.

(17) – Priming, emphasizing word/semantic, and its use in memory tests for patients with Alzheimer’s disease.

(22) – Reading and semantic/word processing tests, especially for patients with Alzheimer's disease.

(21) – Writing and word comprehension, primarily for aphasic subjects, and speech and spelling/grammatical errors secondarily.

Visual/Spatial (75)

(17) – Name and word retrieval, especially of people and objects.

(12) – Semantic categorization for living and non-living objects.

(14) – Confabulation, especially in semantic and episodic memory tasks.

(17) – Agnosia, primarily visual and tactile object recognition and naming disorders.

(15) – Face recognition, emphasizing prosopagnosia, including overt and covert processing of facial identity.

AMNESIA (119)

Amnesia Symptoms (50)

(21) – Retrograde amnesia, with some related emphasis on anterograde amnesia and autobiographical memory.

(14) – Learning, especially in Korsakoff’s amnesia patients with memory problems.

(15) – Amnesia, emphasizing forgetting rates, delays, and recall rates.

Amnesia Physiology (69)

(18) – Temporal lobe problems, especially in patients with temporal lobe epilepsy and/or lobectomy, emphasizing memory impacts, and including hippocampal dysfunction.

(17) – Script generation, mainly in patients with frontal lobe lesions, and associated executive function problems such as sequencing of actions and events for planning towards a goal.

(17) – Memory, emphasizing short-term memory, but including long term as well.

(17) – Spatial orientation/location problems, especially in relation to topographical memory, and the link to lesions, using associated regional cerebral blood flow and MRI measurements.

Summary and Conclusions

Publication Bibliomtrics

The *Cortex* top performers are centered in Italy primarily, and Western Europe generally. The *Neuropsychologia* top performers are centered in the UK primarily, and in the countries of the TTCP (The Technical Cooperation Program) totally. Recent text mining studies in the high tech areas of nanotechnology, nonlinear dynamics, fractals, among others, tend to have a much higher representation from the USA in top performers, some of the Asian countries like Japan, China, and South Korea, and Germany. Except for the USA, none of the top performers from these other countries are represented in these two neuropsychology journals.

It is somewhat surprising that such leading Canadian universities as McGill and the University of Toronto, British universities such as Cambridge University, and American universities such as Harvard University, are not represented in the *Cortex* list of prolific institutions. On the surface, the organizational country distributions appear to be cliquish in nature, with *Cortex’s* prolific institutions being centered around the romance language countries, and *Neuropsychologia’s* being centered around the English-speaking TTCP countries. Both journals might benefit from increased diversification.

Citation Bibliometrics

Even though *Cortex* has reasonable representation from the USA and UK in terms of numbers of authors, they are not getting similar representation in terms of numbers of highly cited papers from these two countries.

These most cited *Cortex* reference documents cover a much longer time domain than is typically found in text mining studies of physical science disciplines. For example, a recent study by the first author on nanotechnology examined a database of articles (from all journals) published in 2003. Of the twenty most highly cited articles, the oldest was 1991, and the median was 1996.

Researchers outside neuropsychology might wonder why more recent papers were not among the list of most cited works in the present study databases. In other fields, the older papers may have only historical interest because the field has moved far beyond the data and theories of that time. The clinical behavioral aspect of neuropsychology emphasized in *Cortex* appears to be different in that regard; change is slow and a method has to be useful and used for many years before it becomes a standard. Owing to the small numbers of patients with interesting syndromes, it is also likely that researchers keep a method long after its initial conception in order not to lose comparison between previous and former/future subjects in experiments. The journals *Neuropsychologia* and *Brain*, which emphasize more the objective diagnostic approaches (as shown in the citation comparison in the text) have slightly more recent average top ten references. This is due to the faster change in the diagnostic technologies used, and the larger share of references devoted to these ‘harder’ technologies.

The most cited articles in *Neuropsychologia* are cited, on average, more than three times as often as the most cited articles in *Cortex*, and the most cited articles in *Brain* are cited, on average, more than twice as often as the most cited articles in *Neuropsychologia*. Whether the difference in highly cited papers is due to the difference in intrinsic quality of the best papers in each journal, the thrust areas selected within the neuropsychology discipline, or the number of people who have access to each journal, or some combination of these causes, cannot be stated at this time.

The average number of citations increases with the average number of authors, ranging from about four authors of the most cited *Cortex* papers to about seven authors of the most cited *Brain* papers. Having more authors may add more dimensions and perspectives to a paper, increasing its comprehensiveness, and having more authors may be the equivalent to having more peer reviewers, increasing the paper’s quality.

The most cited papers have substantially more references than the least cited, in both journals, and the effect is most pronounced in *Neuropsychologia*. Additionally, the average number of citations increases with the average number of references, ranging from about 46 references in the most cited *Cortex* papers to about 68 references in the most cited *Brain* papers. Having more references is one measure of increased scholarship, and may result in additional citations, on average, for historical purposes.

The most significant country representations in the samples examined are the strong positive showing of the UK, and the weak showing of Japan. This latter observation is very different from Japan’s strong showings in almost all of the high tech discipline text mining studies performed by the first author, and suggests that Japanese scientists have not concentrated their energies in the area of neuropsychology.

There is a distinct shift in type of study in proceeding from *Cortex* to *Neuropsychologia* to *Brain*. Clinical behavioral studies, many of them essentially case studies, predominate the most cited *Cortex* papers. *Neuropsychologia* has more of a balance between Behavioral and Diagnostic-Non-Invasive (e.g., PET, MRI) in its ten most cited papers. *Brain* shows a heavy emphasis on Diagnostic-Non-Invasive (7/10), two papers on surgical procedures, and one on Diagnostic-Invasive (e.g., tissue samples). Based on reading Abstracts from each of these journals, the types as represented in the top ten most cited articles roughly approximate the types of papers published overall. Thus, as citations increase in absolute amounts, the study type transitions from the clinically oriented behavioral focus to the correlates with more objective measurements. Also, the references tend to become more recent, on average.

Finally, these bibliometrics results suggest that *Cortex* needs to take steps to attract a more diverse group of highly prolific and cited contributors than its continental Western European base, if it wishes to capture a greater share of the seminal neuropsychology papers. Further investigation of these citation differences is recommended.

Citation Mining

The countries that tend to contribute the most papers to *Cortex* tend also to cite *Cortex* papers most frequently.

Taxonomy

Finally, document clustering showed that the articles sampled in *Cortex* can be reasonably divided into four categories (papers in each category in parenthesis): Semantic Memory (151); Handedness (145); Amnesia (119); and Neglect (66). A similar cluster analysis has not been carried out for *Neuropsychologia* or *Brain*, but the general categories are probably similar, though not necessarily in the same order of importance.

References

Albert ML. Simple test of visual neglect. *Neurology, 23:* 658-664, 1973.

Annett M. *Left, Right, Hand and Brain: The Right Shift Theory.* London: Lawrence Erlbaum 1985.

Axelrod BN and Millis SR. Preliminary standardization of the Cognitive Estimation Test. *Assessment, 1:* 269-274. 1994.

Cutting DR, Karger DR, Pedersen JO and Tukey JW. Scatter/Gather: A cluster-based approach to browsing large document collections. In: *Proceedings of the 15th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'92)*. 318-329, 1992.

Folstein MF, Folstein SE and McHugh PR. Mini-Mental State: A practical method for grading the state of patients for the clinician. *Journal of Psychiatric Research, 12:* 189-198, 1975.

Guha S, Rastogi R and Shim K. CURE: An efficient clustering algorithm for large databases. In: *Proceedings of the ACM-SIGMOD 1998 International Conference on Management of Data (SIGMOD'98)*. 73-84, 1998.

Garfield E. History of citation indexes for chemistry – A brief review. *Journal of Chemical Information and Computer Science, 25:* 170-174, 1985.

Hearst MA. Untangling text data mining. *Proceedings of ACL 99, the 37th Annual Meeting of the Association for Computational Linguistics*, University of Maryland, June 20-26, 1999.

Hearst MA. The use of categories and clusters in information access interfaces. In T Strzalkowski (Ed), *Natural Language Information Retrieval*. Norwell: Kluwer Academic Publishers, 2000.

Karypis G. *CLUTO: A clustering toolkit*. *http://www.cs.umn.edu/~cluto*, 2004.

Karypis G, Haneh and Kumar V. Chameleon: A hierarchical clustering algorithm using dynamic modeling. *IEEE Computer: Special Issue on Data Analysis and Mining, 2:* 68-75. 1999.

Kostoff RN. Database tomography for technical intelligence, *Competitive Intelligence Review, 4:* 38-43, 1993.

Kostoff RN. Database tomography: Origins and applications, *Competitive Intelligence Review, Special Issue on Technology, 5:* 48-55, 1994.

Kostoff RN The use and misuse of citation analysis in research evaluation. *Scientometrics, 43:* 27-43, 1998.

Kostoff RN. Text mining for global technology watch. In M Drake (Ed), *Encyclopedia of Library and Information Science*, Second Edition. New York: Marcel Dekker, 2003, pp. 2789-2799.

Kostoff RN, Buchtel HA, Andrews J and Pfiel KM. The hidden structure of neuropsychology: Text mining of the Journal Cortex: 1991-2001. *DTIC Technical Report Number ADA425 056*, 2004

Kostoff RN, Braun T, Schubert A, Toothman DR and Humenik J. Fullerene roadmaps using bibliometrics and database tomography. *Journal of Chemical Information and Computer Science, 40:* 19-39, 2000a.

Kostoff RN, del Rio JA, Garcia EO, Ramirez AM and Humenik JA. Citation mining: Integrating text mining and bibliometrics for research user profiling. *Journal of the American Society for Information Science and Technology, 52:* 1148-1156, 2001.

Kostoff RN and Demarco RA. Science and technology text mining. *Analytical Chemistry, 73:* 370-378A, 2001.

Kostoff RN, Eberhart HJ and Miles D. System and Method for Database Tomography, U.S. Patent Number 5440481, August 8, 1995.

Kostoff RN, Eberhart HJ and Toothman DR. Database tomography for information retrieval, *Journal of Information Science, 23:* 301-311, 1997a.

Kostoff RN, Eberhart HJ and Toothman DR. Database tomography for technical intelligence: A roadmap of the near-earth space science and technology literature. *Information Processing and Management, 34:* 69-85, 1998.

Kostoff RN, Eberhart HJ and Toothman DR. Hypersonic and supersonic flow roadmaps using bibliometrics and database tomography. *Journal of the American Society for Information Science, 50:* 427-447, 1999.

Kostoff RN, Eberhart HJ, Toothman DR and Pellenberg R. Database tomography for technical intelligence: Comparative roadmaps of the research impact assessment literature and the Journal of the American Chemical Society, *Scientometrics, 40:* 103-134, 1997b.

Kostoff RN and Geisler E. Strategic management and implementation of textual data mining in government organizations. *Technology Analysis and Strategic Management, 11:* 493-525, 1999.

Kostoff RN, Green KA, Toothman DR and Humenik J. Database tomography applied to an aircraft science and technology investment strategy. *Journal of Aircraft, 37:* 727-730, 2000b.

Lostewicz P, Oard D and Kostoff RN. Textual data mining to support science and technology management. *Journal of Intelligent Information Systems, 15:* 99-119, 2000.

Mac Roberts M and Mac Roberts B. Problems of citation analysis. *Scientometrics, 36:* 435-444, 1996.

McKhann G, Drachman D, Folstein M, Katzman R, Price D and Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. *Neurology, 34:* 939-944, 1984.

Nelson HE. A modified card sorting test sensitive to frontal lobe defects. *Cortex, 12:* 313-324, 1976.

Oldfield RC. The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia, 9:* 97–113, 1971.

Prechelt L, Malpohl G and Pilippsen M. Finding plagiarisms among a set of programs with JPlag. *Journal of Universal Computer Science, 8:* 1016-1038, 2002.

Rasmussen E. Clustering Algorithms. In WB Frakes and R Baeza-Yates (Eds), *Information Retrieval Data Structures and Algorithms*. Upper Saddle River: Prentice Hall, 1992.

Science Citation Index. *ISI Web of Science*. Thomson ISI. Philadelphia, 2004.

Shallice T. *From Neuropsychology to Mental Structure*. Cambridge: Cambridge University Press 1988.

Shallice T and Evans ME. The involvement of the frontal lobes in cognitive estimation. *Cortex, 14:* 294-303,1978.

Snodgrass JG and Vanderwart M. A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6:* 174-215, 1980.

Steinbach M, Karypis G and Kumar V. A comparison of document clustering techniques. *Technical Report N. 00-034. 2000*. Department of Computer Science and Engineering. University of Minnesota, 2000.

Talairach J and Tournoux P. *Co-planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System – an Approach to Cerebral Imaging*. New York: Thieme Medical Publishers, 1988.

Willet P. Recent trends in hierarchical document clustering: A critical review. *Information Processing and Management, 24:* 577-597, 1988.

Wise MJ. String similarity via greedy string tiling and running Karb-Rabin matching. ftp://ftp.cs.su.oz.au/michaelw/doc/RKR_GST.ps, 1992. Dept. of CS, University of Sidney.

Zamir O and Etzioni O. Web document clustering: A feasibility demonstration. In: *Proceedings of the 19th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'98)*, 46-54, 1998.

Zhao Y and Karypis G. Criterion functions for document clustering: Experiments and analysis. *Machine Learning*, in press.

Zhu DH and Porter AL. Automated extraction and visualization of information for technological intelligence and forecasting. *Technological Forecasting and Social Change, 69:* 495-506 2002.

Ronald N. Kostoff, Ph.D., Office of Naval Research, 800 N. Quincy St., Arlington, VA 22217, USA.