Jointly published by Elsevier Science Ltd, Oxford and Akadémiai Kiadó, Budapest

Scientometrics, Vol. 38, No. 1 (1997) 175–204

DIMENSIONS OF SCIENTOMETRIC INDICATOR DATAFILES
WORLD SCIENCE IN 1990-1994

T. BRAUN, A. SCHUBERT

Information and Scientometrics Research Unit (ISSRU), Library of the Hungarian Academy of Sciences, H-1245, P.O.Box 1002, Budapest, Hungary
E-mail: h1533bra@ella.hu

(Received October 3, 1996)

Scientometric indicators are treated according to dimensional approaches. One, two, three, dimensions and multidimensional characteristics are revealed graphically for giving a panoramic view on the publication activity and citation impact of different countries.

Introduction

The significance of the concept of *multidimensionality* in scientometrics was passionately advocated by *Moravcsik*,[1, 2] and his views have been widely echoed ever since (e.g., Ref. 3-5). In presenting data from ISSRU's *Scientometric Indicators Datafiles*,[6] due stress has always been laid to this important aspect. In the present study, the full scale of one- to several-dimensional presentation techniques will be used to display the data of the most recent five-year period, 1990-1994, of the *Datafiles*. We should also like to take the opportunity to reiterate the most conspicuous general conclusions that can be drawn from representations in various dimensions.

Data sources and sata processing

The basic data source of the *Datafiles* is the *Science Citation Index* (*SCI*)® database of the Institute for Scientific Information (ISI, Philadelphia, PA, USA). Design and realization of the computer software for building and processing the datafiles is the product of ISSRU, Budapest.

0138–9130/97/US $ 15.00

Scientometric indicators

Most of the details concerning methodology, terminology and definitions are exactly the same as in our earlier publications (e.g., Ref. 7-9). Only the most fundamental points are reiterated here very briefly.

Subject classification and country assignment. Papers were classified into science fields on the basis of the field classification of the publishing journals. Subfields used by the *SCI* database were grouped into five major fields: Life Sciences, Physics, Chemistry, Engineering amd Mathematics. Country assignment was based on first corporate address (usually the address of the first author). 44 countries of considerable publication output in the period 1990-1994 were selected to be presented in this study. Countries are usually denoted by their triliteral ISO standard codes (see *Appendix 1*).

Publication and citation counts. Publications included in the 1990-1994 annual cumulations of the CD-ROM Edition of the *SCI* database as articles, letters, notes and reviews were considered; all citations received by these items in the same period were counted.

Activity Index (AI). The Activity Index is defined as AI = (the country's world share in publications of a given field)/(the country's world share in publications of all science fields combined).

AI characterizes the relative research effort a country devotes to a given field. AI = 1 indicates that the country's research effort in the given field corresponds to the world average. AI < 1 reflects lower-than-average effort, AI > 1 reflects higher-than-average effort. It must be stressed that AI characterizes the relative distribution of a country's efforts within the fields, therefore, a country's activity cannot be above the average in all fields: higher-than-average fields must necessarily be balanced by lower-than average ones.

Attractivity Index (AAI). The Attractivity Index is defined as AAI = (the country's world share in citations to a given field)/(the country's world share in citations to all science fields combined).

AAI characterizes the relative impact of a country's publications in a given field as reflected in the citations they attract. AAI = 1 indicates that the country's impact in the given field corresponds to the world average. AAI < 1 reflects lower-than-average impact, AI > 1 reflects higher-than-average impact. It must be stressed that AAI characterizes the relative distribution of a country's impact within the fields, therefore, a country's attractivity cannot be above the average in all fields: higher-than-average fields must necessarily be balanced by lower-than average ones.

Mean Observed Citation Rate (MOCR). Average citation rate per publication, i.e., (number of citations)/(number of publications).

Mean Expected Citation Rate (MECR). Average expected citation rate per publication, i.e., (expected number of citations)/(number of publications), where the expected number of citations is calculated on the basis of the average citation rates of the publishing journal, i.e., each paper is expected to receive the citation rate of an average paper of the same age in the same journal.

Relative Citation Rate. Relative Citation Rate is the ratio of Mean Observed Citation Rate (MOCR) to Mean Expected Citation Rate (MECR). RCR = 1 indicates that the citation average of the country exactly matches the expected value in this field. RCR > 1 (< 1) reflects a citation average above (below) the expectations.

One-dimensional representations

The simplest and, in a sense, most direct presentation of scientometric indicators is in the form of ranked lists. The question is, of course, by which indicator should the ranking be based. Ranking by absolute numbers of publications, citations, etc. trivially results in the scientific "superpowers" on the top, specific variations appear only as small differences of large numbers. By using well chosen relative indicators, the specific eminence of small or medium-sized countries in certain fields can be made evident. Relative Citation Rate (RCR) has proven to be rather suitable for this purpose. In Table 1 the top section of the lists of countries ranked by RCR is given in each of the 5 major science fields, as well as in all science fields combined. Total number of publication is given as supplementary information.

By emphasizing the above-average performances (RCR ≥ 1), we should like to call attention to the *principle of antidiagnostics*: while in medical diagnosis numerical laboratory results can indicate only pathological status but not health, in scientometrics, numerical indicators can reliably suggest only eminence but never worthlessness. The level of citedness, for instance, may be effected by numerous factors other than inherent scientific merits, but without such merits no statistically significant eminence in citedness can be achieved.

The main moral of the results presented in Table 1 is that "small is beautiful": some small or medium-sized developed countries (typically the Scandinavian and the Benelux countries and Switzerland) occupy, as a rule, the first positions, followed by the "big four" (USA, UK, Germany and France) and some ambitious "outsiders". As compared to our earlier similar studies, the emergence of Ireland, South Africa and Singapore is striking.

Table 1
Top ranked countries by Relative Citation Rate (RCR)

All science fields combined

Rank	Code	RCR	# publ.
1	CHE	1.16	32977
2	DNK	1.11	20002
3	NLD	1.10	53435
4	SWE	1.10	41629
5	DEU	1.09	170007
6	UKD	1.08	217504
7	USA	1.06	893886
8	FIN	1.06	17113
9	IRL	1.00	4964
10	BEL	1.00	22487
11	FRA	1.00	130679

Life Sciences

Rank	Code	RCR	# publ.
1	DNK	1.13	15035
2	SWE	1.11	30387
3	CHE	1.11	19144
4	UKD	1.10	145799
5	NLD	1.10	35799
6	FIN	1.09	12570
7	DEU	1.06	86269
8	USA	1.05	559463
9	NZL	1.03	8300
10	BEL	1.02	14551
11	NOR	1.00	8715

Physics

Rank	Code	RCR	# publ.
1	CHE	1.27	7644
2	DEU	1.12	41097
3	NLD	1.11	8817
4	UKD	1.10	33012
5	USA	1.09	165990
6	DNK	1.08	2813
7	FRA	1.05	29863
8	IRL	1.04	782
9	SWE	1.01	5611

Table 1 (cont.)

Chemistry

Rank	Code	RCR	# publ.
1	CHE	1.20	4238
2	NLD	1.19	5359
3	SWE	1.11	3599
4	SGP	1.10	559
5	USA	1.09	77329
6	DNK	1.08	1273
7	UKD	1.06	20114
8	DEU	1.05	28210
9	AUT	1.04	1755
10	CAN	1.01	10175
11	NOR	1.00	924

Engineering

Rank	Code	RCR	# publ.
1	IRL	1.57	492
2	DNK	1.27	1352
3	CHE	1.22	2998
4	FIN	1.21	1590
5	NLD	1.16	5033
6	DEU	1.14	19337
7	USA	1.12	99203
8	SWE	1.10	3161
9	UKD	1.10	20275
10	BEL	1.09	1970
11	FRA	1.07	12814
12	AUS	1.03	4649
13	AUT	1.00	1519

Mathematics

Rank	Code	RCR	# publ.
1	DNK	1.79	271
2	ZAF	1.19	166
3	UKD	1.19	3448
4	FIN	1.09	259
5	USA	1.08	20989
6	NOR	1.08	240
7	NLD	1.06	995
8	FRA	1.02	2713
9	SGP	1.01	139
10	EGY	1.01	129

Two-dimensional representations

The most frequently proclaimed objection against the use of RCR as size- and research field-independent indicator of citation impact is that it might unjustly favour countries regularly publishing in low-impact journals, the low standards of which is easy to surpass. It has long been asserted, however, that this is not the case. On the contrary, most citations in lower impact journals are attracted by authors, institutions and, thereby, countries regularly publishing in high prestige journals, where they could become known and gain recognition.

A rather convincing evidence supporting this relation is the typical shape of the MOCR vs. MECR diagrams (the so-called "relational charts", cf. Ref. 7). The diagrams of Fig. 1 have a characteristic upward bend. The upper left fields of the charts remain practically empty; those, who typically publish in low impact journals, cannot reach even their moderate expected citation level. (*Bonitz*[10] relates this phenomenon to Merton's celebrated Matthew-effect.)

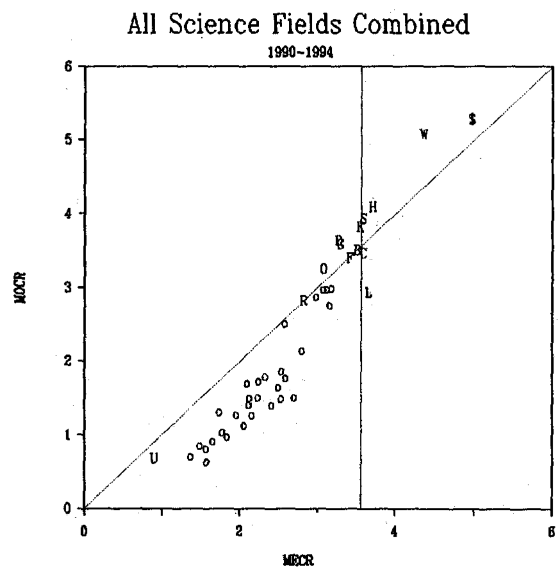
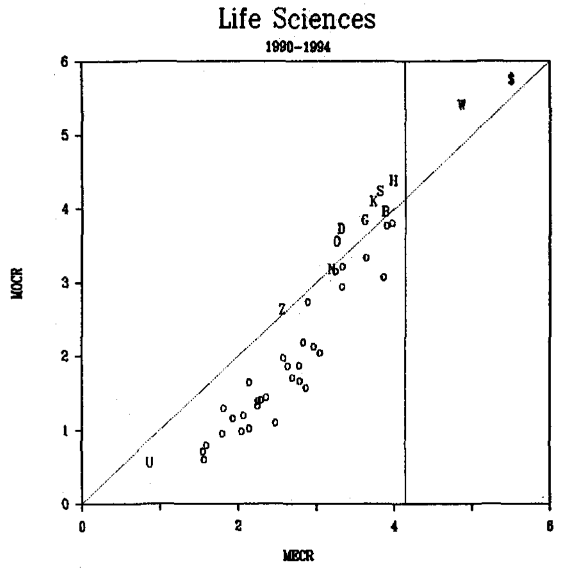


Fig. 1. Relational charts of observed vs. expected citation rates. $: USA, B: BEL, C: CAN, D: DNK, F: FRA, G: DEU, H: NLD, K: UKD, L: ISR, O: FIN, R: IRL, S: SWE, U: SUN, W: CHE



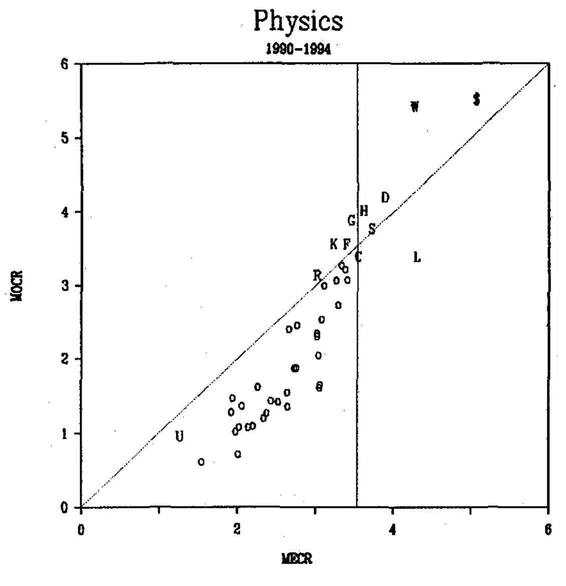
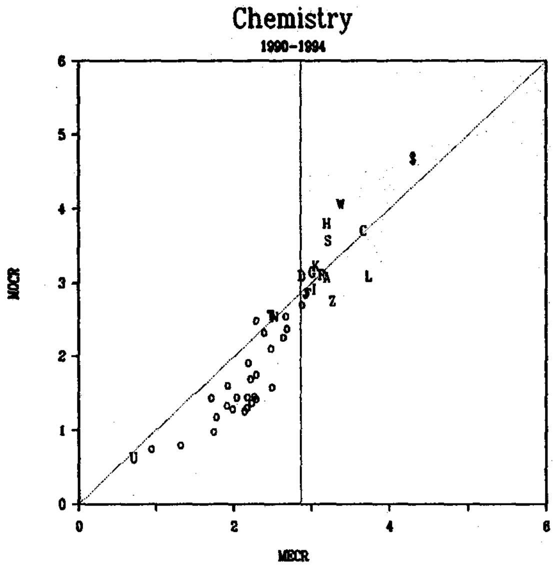


Fig. 1. (cont.). $: USA, B: BEL, C: CAN, D: DNK, F: FRA, G: DEU, H: NLD, K: UKD, L: ISR, O: FIN, R: IRL, S: SWE, U: SUN, W: CHE



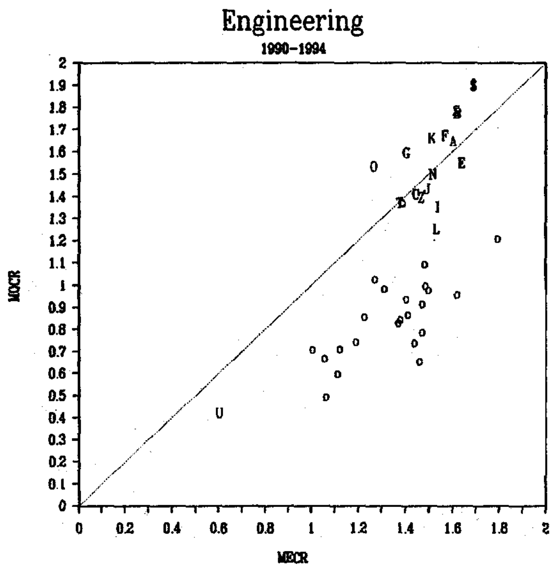


Fig. 1. (cont.). $: USA, B: BEL, C: CAN, D: DNK, F: FRA, G: DEU, H: NLD, K: UKD, L: ISR, O: FIN, R: IRL, S: SWE, U: SUN, W: CHE

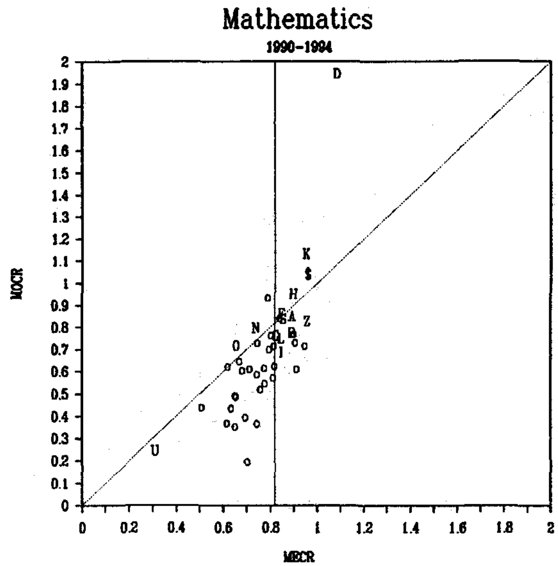


Fig. 1. (cont.). $: USA, B: BEL, C: CAN, D: DNK, F: FRA, G: DEU, H: NLD, K: UKD, L: ISR, O: FIN, R: IRL, S: SWE, U: SUN, W: CHE

In conform with the "antidiagnostic principle" introduced above, only those countries are identified on the diagrams, which are, in some sense, eminent. A rather characteristic position is that of Israel (denoted by the letter L on the charts) with very ambitious publication strategy (publishes in the highest impact journals) but with relatively modest RCR. The advantage of the "two-dimensional" thinking is that we need not decide whether this strategy worth more or less than another country's average citation rate on a much lower expected level; we can record the differences and wonder on the causes and effects.

Three-dimensional representations

On the relational charts of Fig. 1 all countries look alike (except for the individual labelling of the eminents). By adding a third dimension to the diagram, an idea of the size of the single countries can also be given. The "scientometric landscape" shown in Fig. 2 is rather similar to those published earlier.[11]

The most striking feature is the two solitary peaks at the two opposite ends of the "sierra": the USA and the (former) USSR. Conspicuously, both are quite close to the main diagonal: their actual citation rate is close to the expected level. The difference is in the actual level of the citation rates and the rationale behind. In the US, as the saying goes, for any given thing there is another one equal in magnitude and opposite in sign,

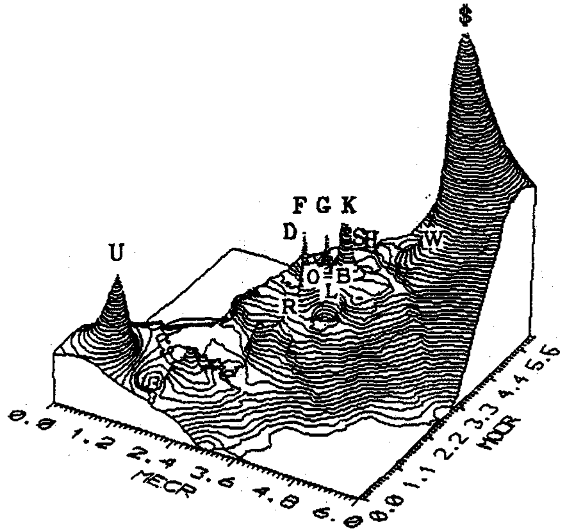


Fig. 2. Landscape of observed vs. expected citation rates. (All science fields combined, 1990-1994)
$: USA, B: BEL, C: CAN, D: DNK, F: FRA, G: DEU, H: NLD, K: UKD, L: ISR, O: FIN, R: IRL, S: SWE, U: SUN, W: CHE

that is, a huge amount of highly cited publications is counterbalanced by an equally huge amount of poorly cited or uncited items. In the USSR, because of the notoriously closed publication policy (at least as the past is concerned), the citation rates are practically gauged against themselves.

Multi-dimensional representations

The whys and hows of multidimensional data representation in statistics is widely discussed in the literature and, most recently, also on the Net.[12] Most of these techniques have existed for many years prior to the advent of statistical computing. However, computing power has led to new possibilities that requires reassessing previous practice. Some of the methods are included in such widespread commercial softwares as Excel or Mathematica, thereby hardly any time, energy and money is to be spent to experiment with dozens of various options to find an optimal representation.

Radar plot, a standard feature in Excel's Gallery, appeared very suitable to compare the distribution of activity over the five major science fields. In Figure 3, countries are grouped by their Activity Indexes: how many and which AI's are above unity.

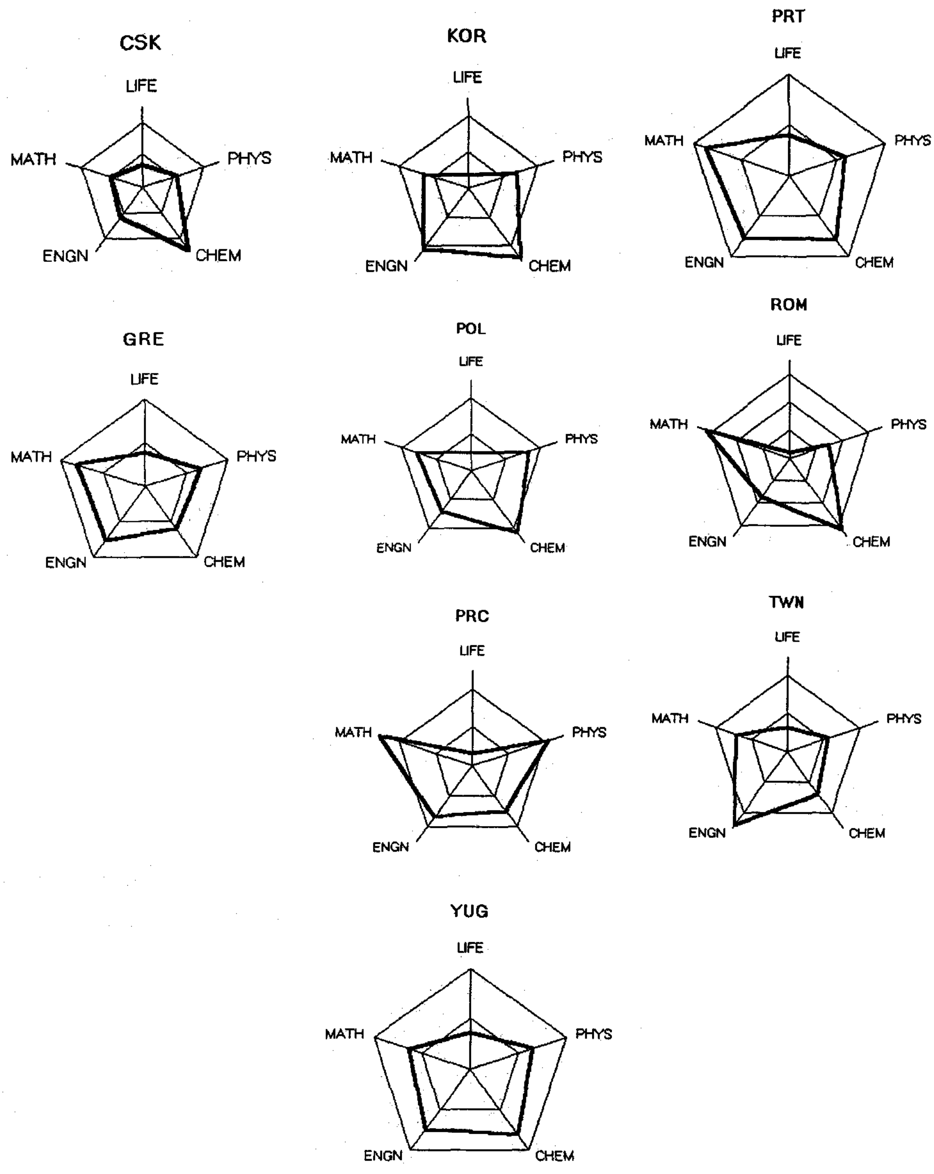


Fig. 3a. Radar plots of Activity Indexes. AI>1 for Physics, Chemistry, Engineering and Mathematics

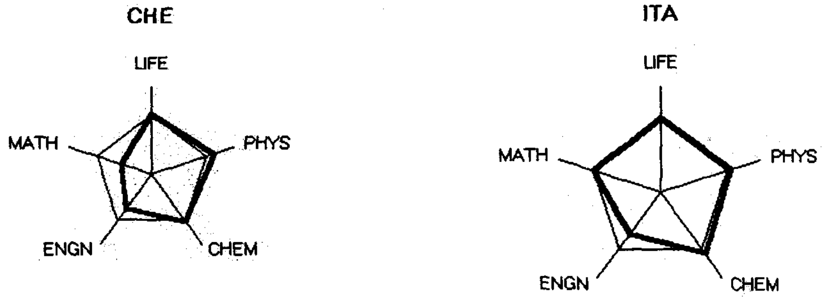


Fig. 3b. Radar plots of Activity Indexes. AI>1 for Life Sciences, Physics and Chemistry

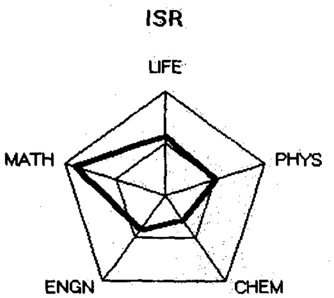


Fig. 3c. Radar plots of AIs. AI>1 for Life Sciences, Physics and Mathematics

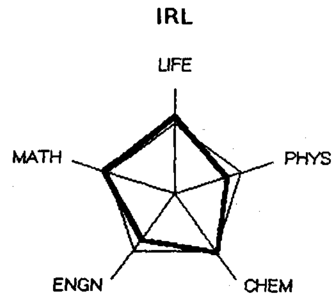


Fig. 3d. Radar plots of AIs. AI>1 for Life Sciences, Chemistry and Mathematics

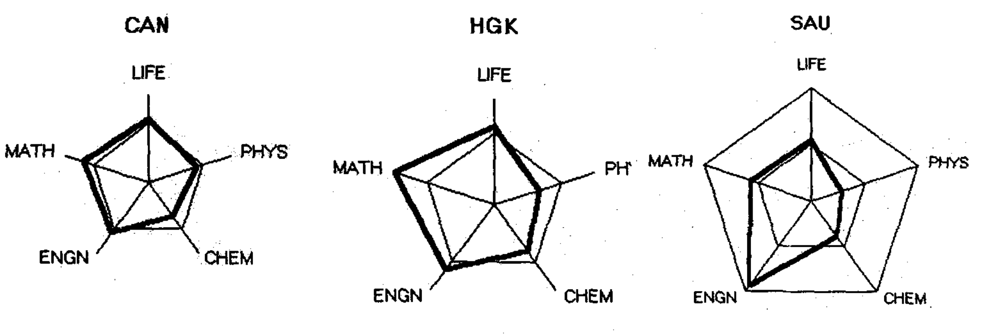


Fig. 3e. Radar plots of Activity Indexes. AI>1 for Life Sciences, Engineering and Mathematics

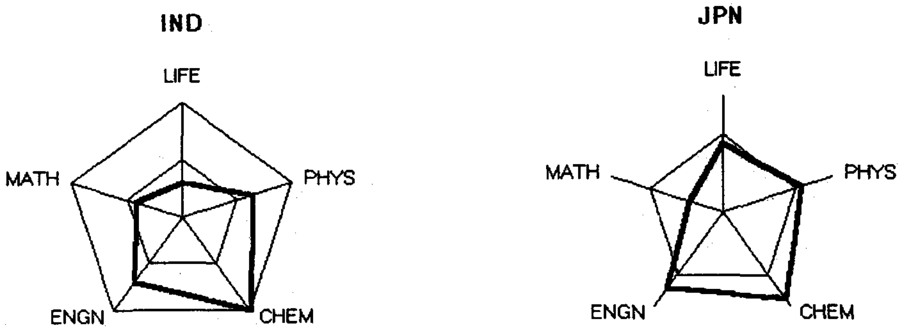


Fig. 3f. Radar plots of Activity Indexes. AI>1 for Physics, Chemistry and Engineering

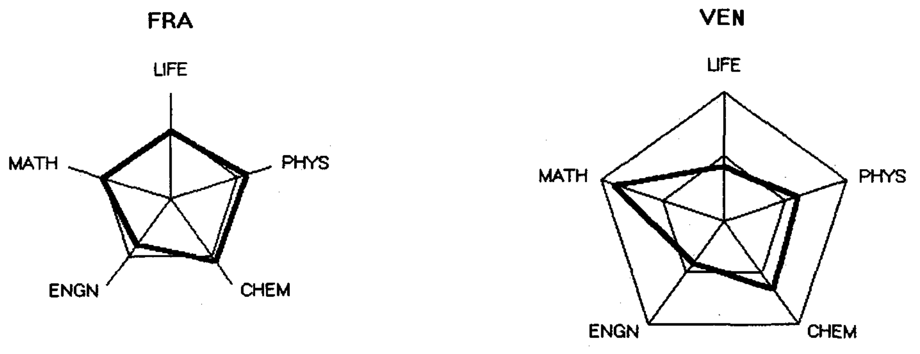


Fig. 3g. Radar plots of Activity Indexes. AI>1 for Physics, Chemistry and Mathematics

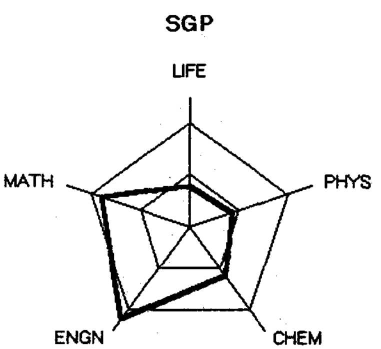


Fig. 3h. Radar plots of Activity Indexes. AI>1 for Chemistry, Engineering and Mathematics

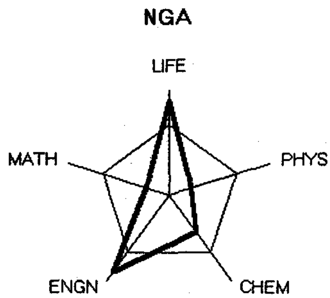


Fig. 3i. Radar plots of Activity Indexes. AI>1 for Life Sciences and Engineering

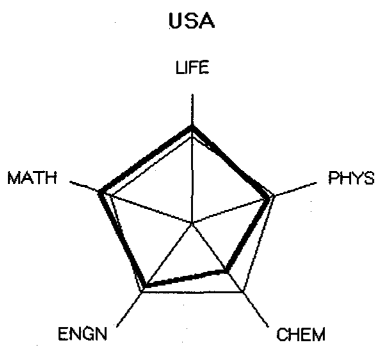


Fig. 3j. Radar plots of Activity Indexes. AI>1 for Life Sciences and Mathematics

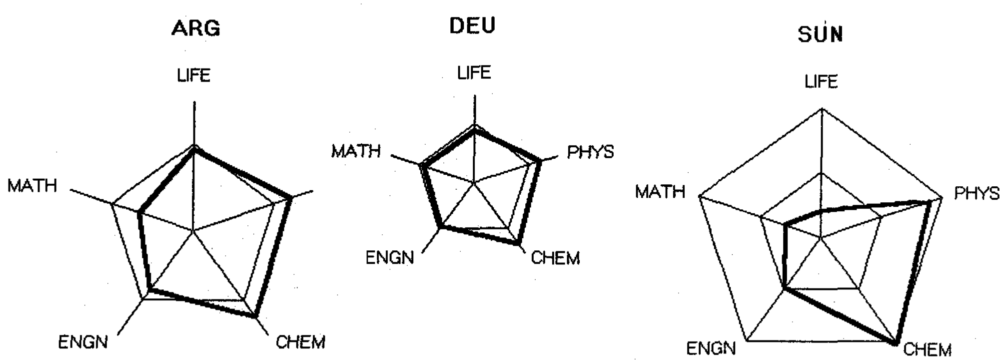


Fig. 3k. Radar plots of Activity Indexes. AI>1 for Physics and Chemistry

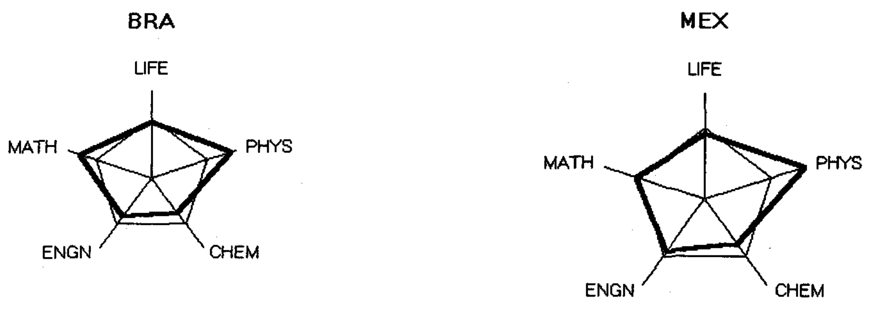


Fig. 3l. Radar plots of Activity Indexes. AI>1 for Physics and Mathematics

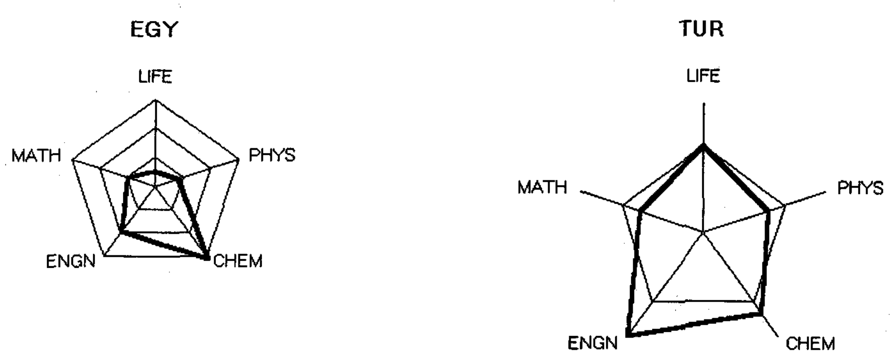


Fig. 3m. Radar plots of Activity Indexes. AI>1 for Chemistry and Engineering

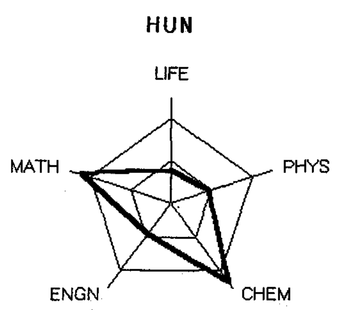


Fig. 3n. Radar plots of AIs. AI>1 for Chemistry and Mathematics

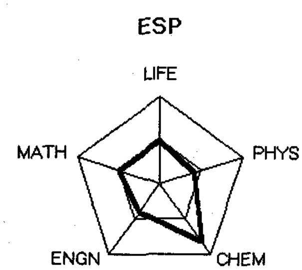


Fig. 3o. Radar plots of AIs. AI>1 for Chemistry

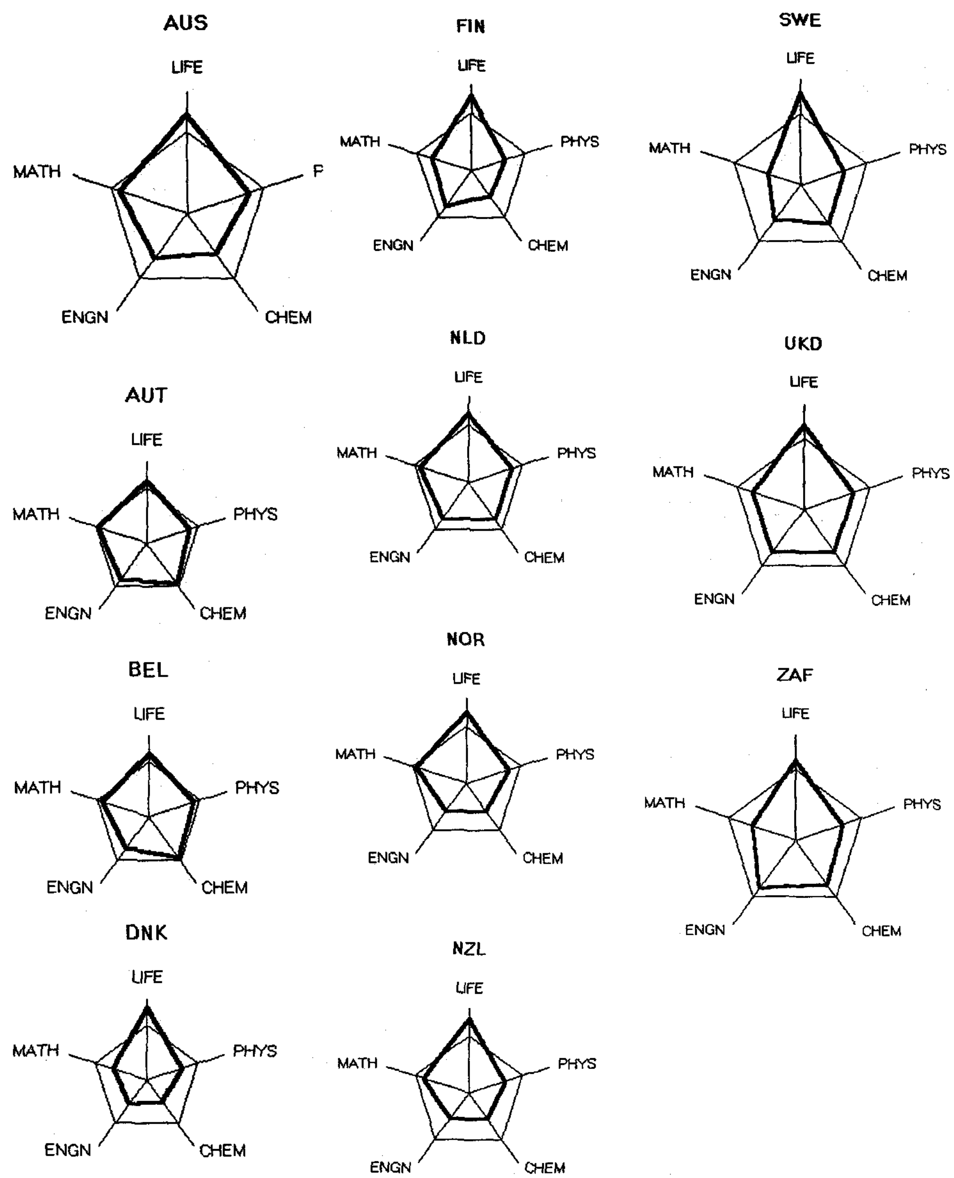


Fig. 3p. Radar plots of Activity Indexes. AI>1 for Life Sciences

Another spectacular technique of visualizing multidimensional statistical data is that of Chernoff faces. In an earlier paper,[5] it has already been used to represent scientometric indicators of developing countries. In the picture gallery of Fig. 5, features of the whole set of 44 countries can be surveyed in the five major science fields. The length of the face represents the Activity Index, the size of the eyes the Attractivity Index, the length of the nose the Mean Expected Citation Rate and the smile the Relative Citation Rate. For comparison, in Fig. 4 the "standard face", i.e., that corresponding to world average data is given.

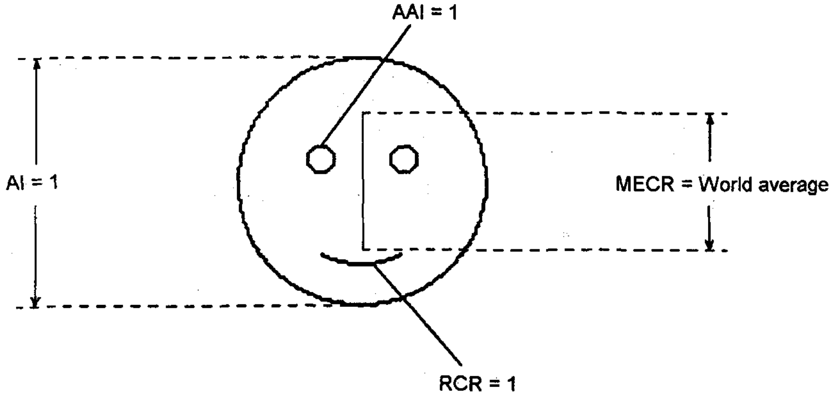


Fig. 4.Chernoff faces with features corresponding to world averages

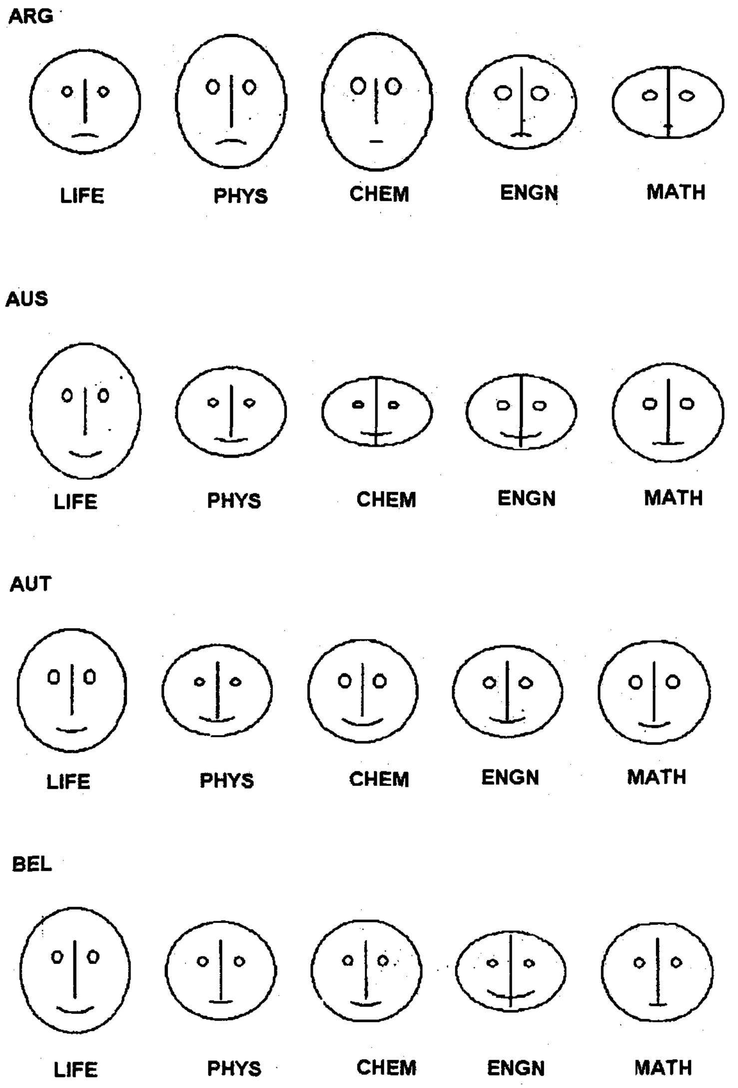


Fig. 5. Chernoff faces representing the main scientometric indicators

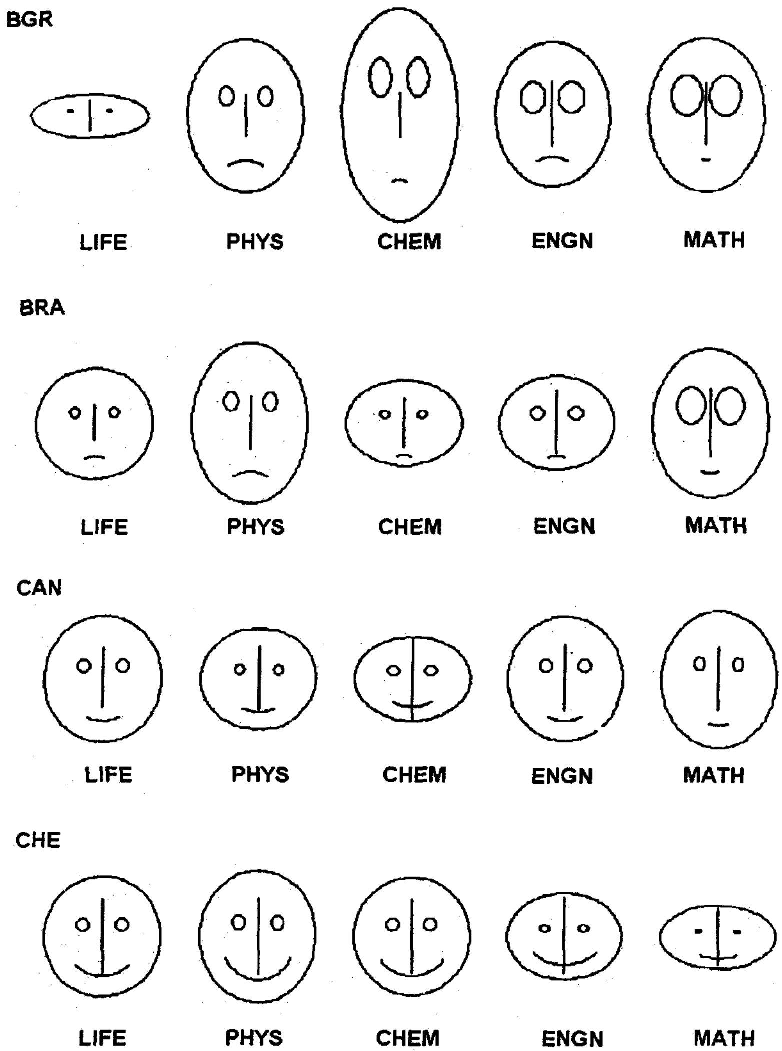

BGR
LIFE
PHYS
CHEM
ENGN
MATH
BRA
LIFE
PHYS
CHEM
ENGN
MATH
CAN
LIFE
PHYS
CHEM
ENGN
MATH
CHE
LIFE
PHYS
CHEM
ENGN
MATH

Fig. 5. (cont.)

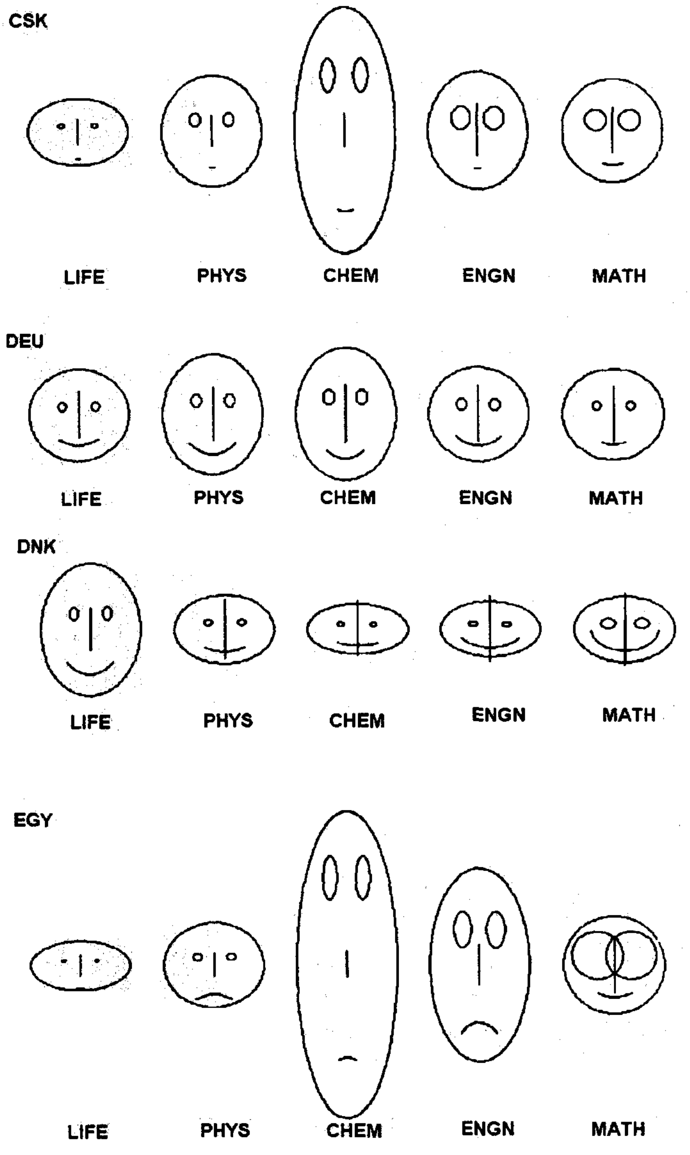

CSK
LIFE
PHYS
CHEM
ENGN
MATH
DEU
LIFE
PHYS
CHEM
ENGN
MATH
DNK
LIFE
PHYS
CHEM
ENGN
MATH
EGY
LIFE
PHYS
CHEM
ENGN
MATH

Fig. 5. (cont.)

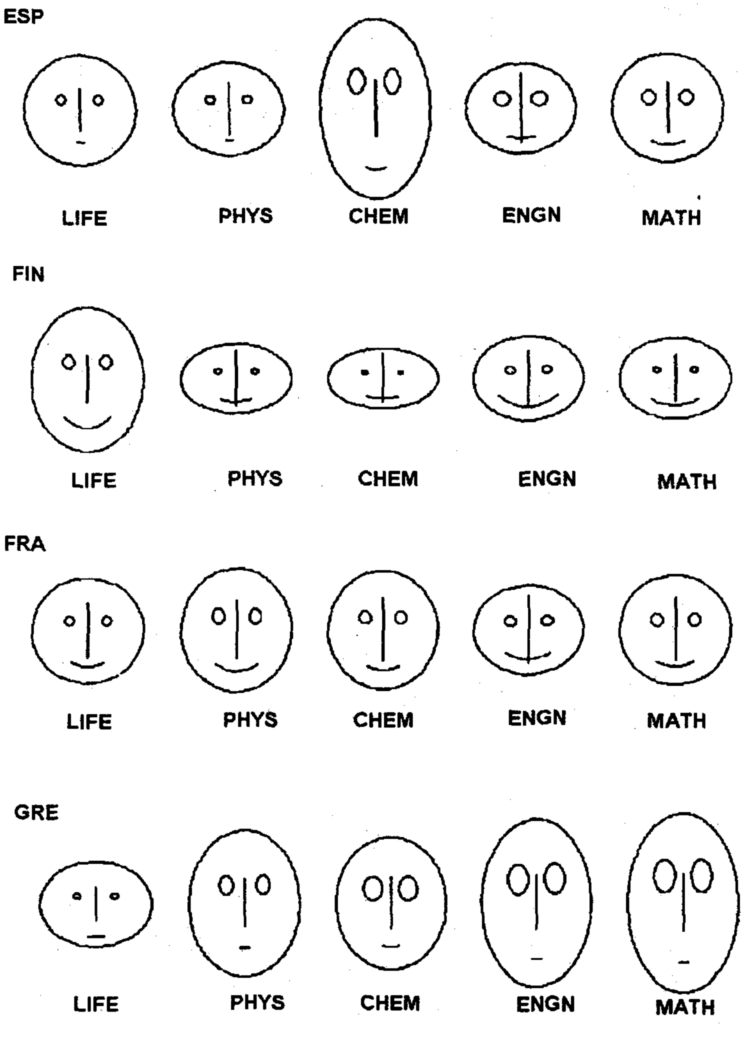

ESP
LIFE
PHYS
CHEM
ENGN
MATH
FIN
LIFE
PHYS
CHEM
ENGN
MATH
FRA
LIFE
PHYS
CHEM
ENGN
MATH
GRE
LIFE
PHYS
CHEM
ENGN
MATH

Fig. 5. (cont.)

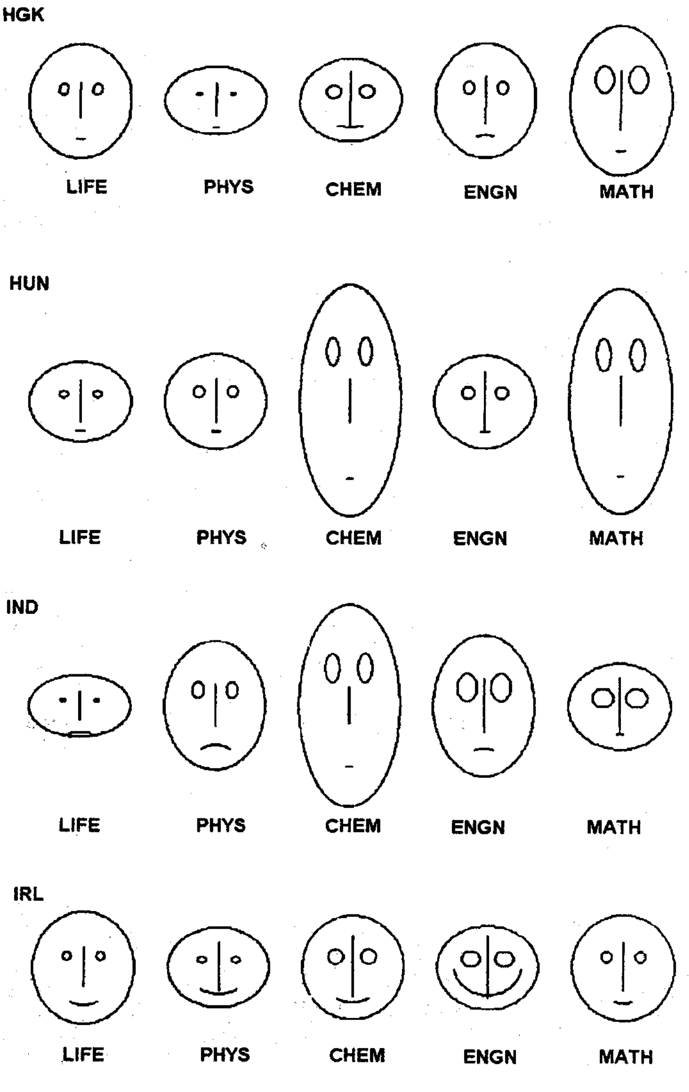

HGK
LIFE
PHYS
CHEM
ENGN
MATH
HUN
LIFE
PHYS
CHEM
ENGN
MATH
IND
LIFE
PHYS
CHEM
ENGN
MATH
IRL
LIFE
PHYS
CHEM
ENGN
MATH

Fig. 5. (cont.)

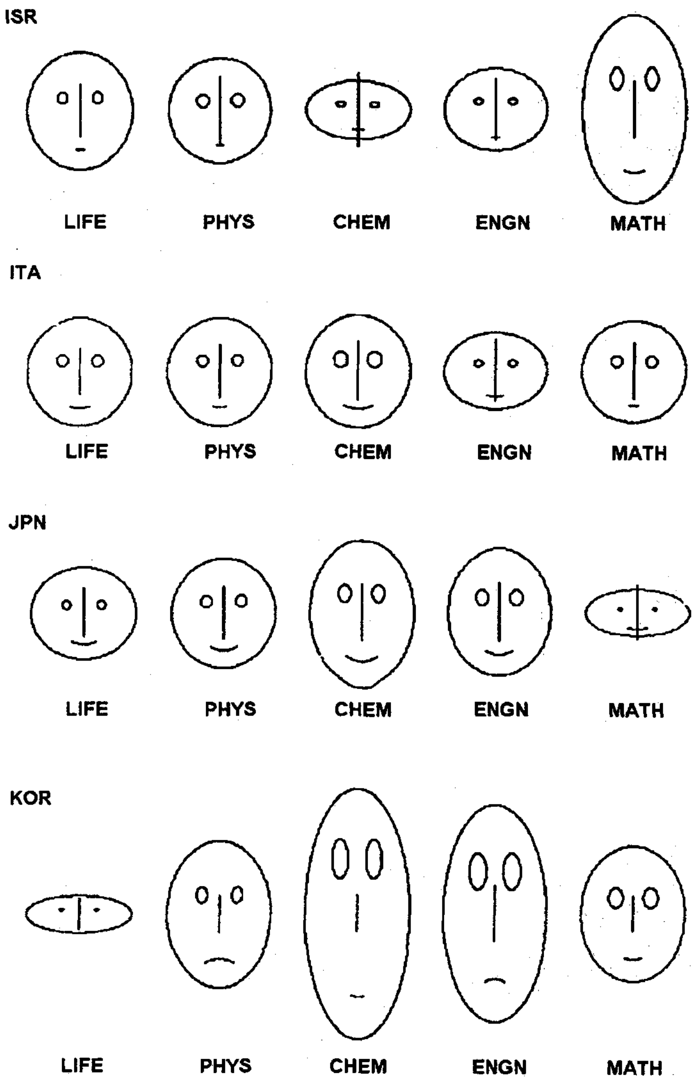

ISR
LIFE
PHYS
CHEM
ENGN
MATH
ITA
LIFE
PHYS
CHEM
ENGN
MATH
JPN
LIFE
PHYS
CHEM
ENGN
MATH
KOR
LIFE
PHYS
CHEM
ENGN
MATH

Fig. 5. (cont.)

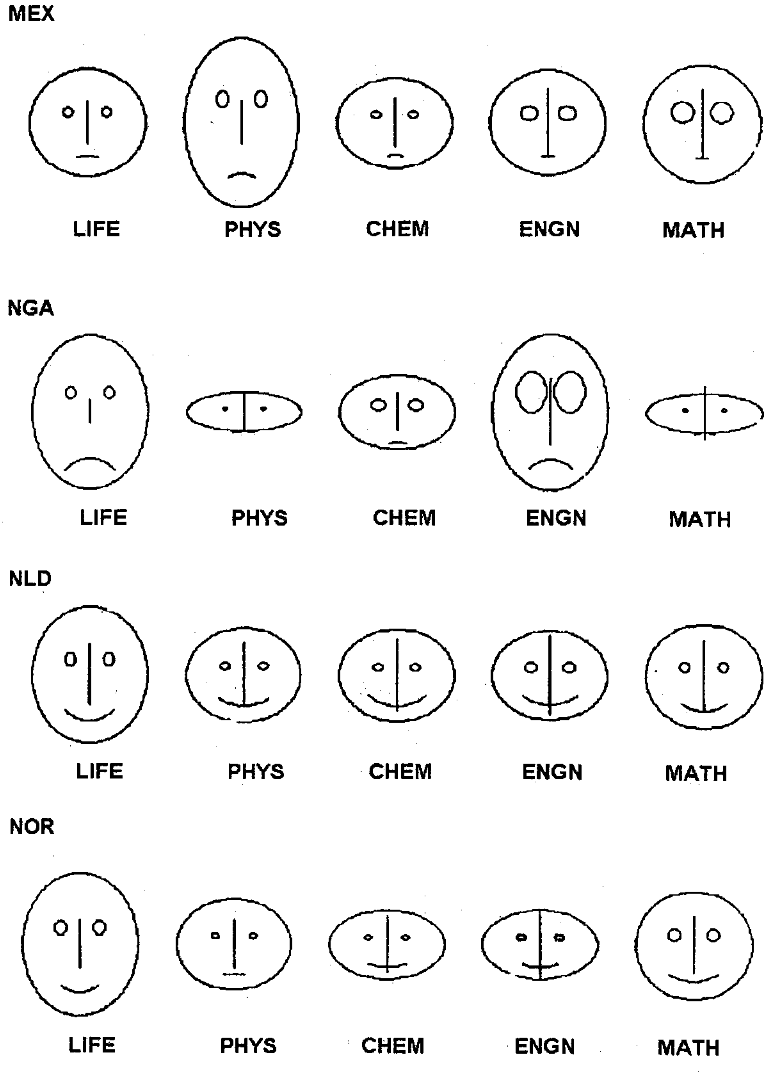

MEX
LIFE
PHYS
CHEM
ENGN
MATH
NGA
LIFE
PHYS
CHEM
ENGN
MATH
NLD
LIFE
PHYS
CHEM
ENGN
MATH
NOR
LIFE
PHYS
CHEM
ENGN
MATH

Fig. 5. (cont.)

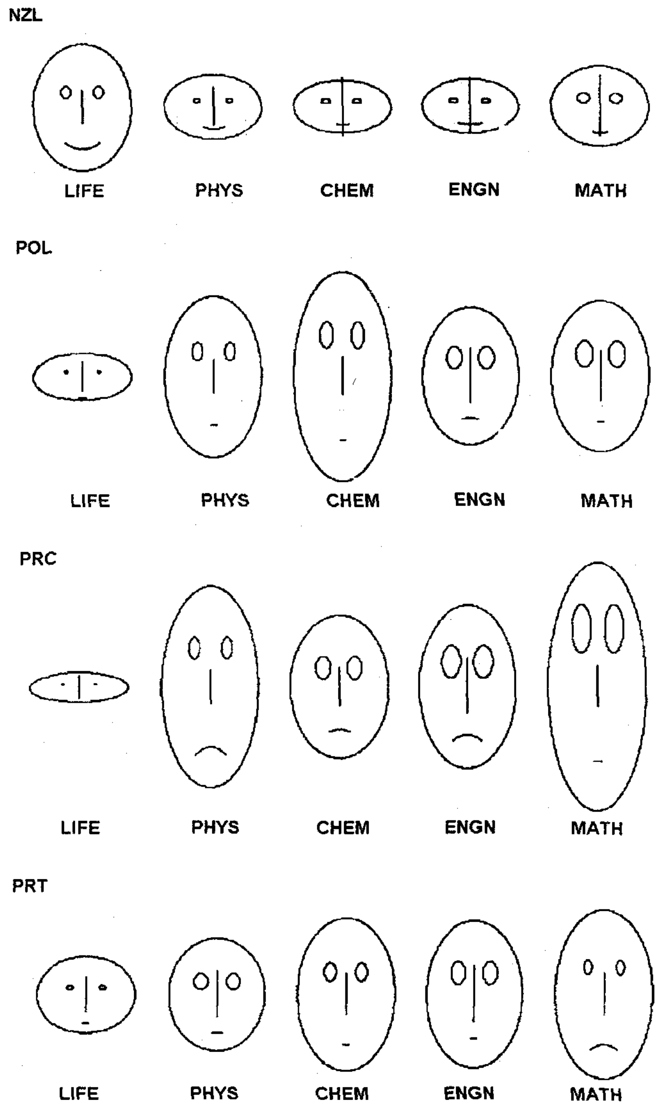

NZL
LIFE
PHYS
CHEM
ENGN
MATH
POL
LIFE
PHYS
CHEM
ENGN
MATH
PRC
LIFE
PHYS
CHEM
ENGN
MATH
PRT
LIFE
PHYS
CHEM
ENGN
MATH

Fig. 5. (cont.)

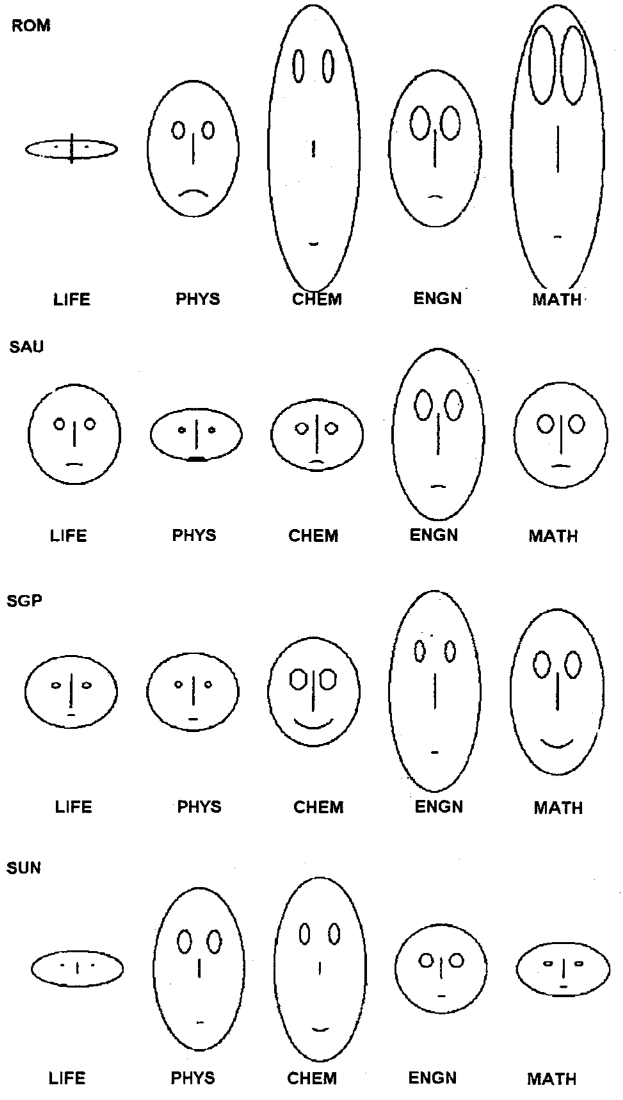

ROM
LIFE
PHYS
CHEM
ENGN
MATH
SAU
LIFE
PHYS
CHEM
ENGN
MATH
SGP
LIFE
PHYS
CHEM
ENGN
MATH
SUN
LIFE
PHYS
CHEM
ENGN
MATH

Fig. 5. (cont.)

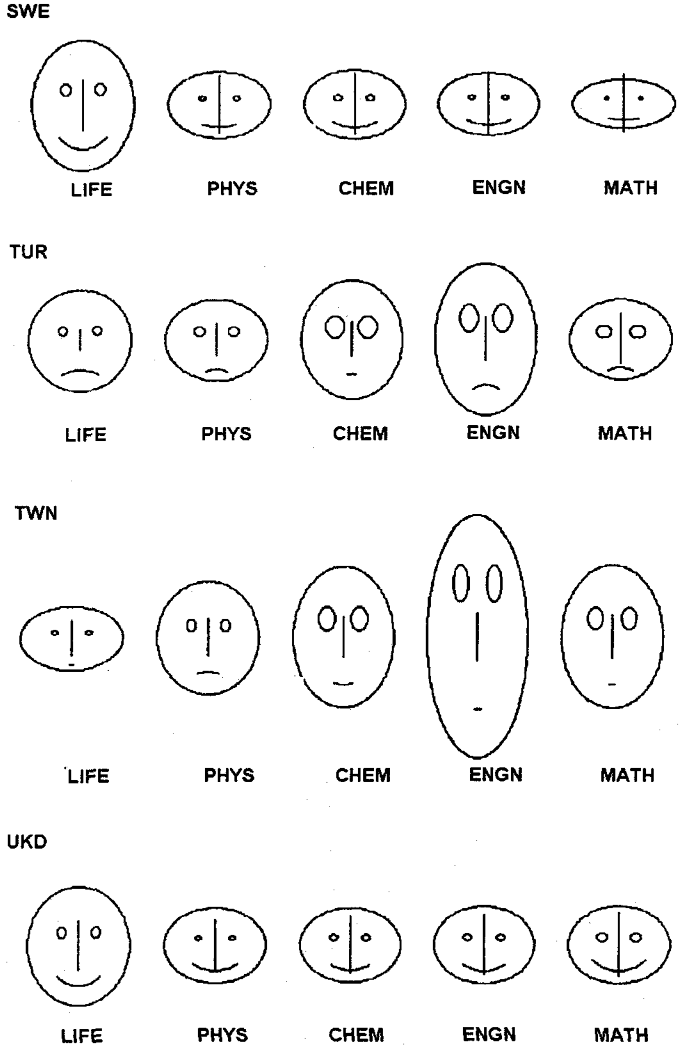

SWE
LIFE
PHYS
CHEM
ENGN
MATH
TUR
LIFE
PHYS
CHEM
ENGN
MATH
TWN
LIFE
PHYS
CHEM
ENGN
MATH
UKD
LIFE
PHYS
CHEM
ENGN
MATH

Fig. 5. (cont.)

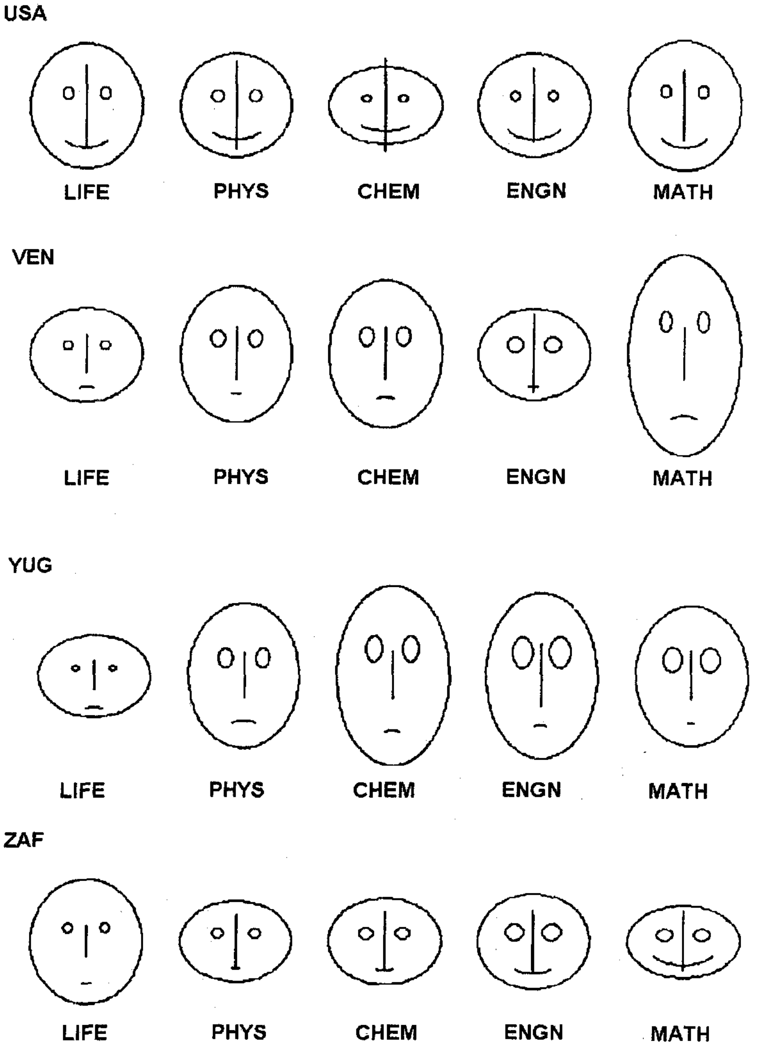

USA
LIFE
PHYS
CHEM
ENGN
MATH
VEN
LIFE
PHYS
CHEM
ENGN
MATH
YUG
LIFE
PHYS
CHEM
ENGN
MATH
ZAF
LIFE
PHYS
CHEM
ENGN
MATH

Fig. 5. (cont.)

Conclusions

With the examples given in this study, we wanted to illustrate that one-, two-, three- and multidimensional representations all have their specific role in the comprehensive and careful analysis of scientometric indicators. However important the choice of dimensionality and form of representation might be, it should always be remembered that this is only one phase of a complex process, which begins with the conception and construction of a system of indicators and ends with their purposeful and fair use. Both ends of the process are crucial; as to the first, thoughtfully built indicator databases (such as ISSRU's Scientometric Indicator Datafiles is attempted to be) are required, as to the second, well-conceived general guiding principles (such as the "principle of antidiagnosis") might be helpful.

*

The authors gratefully acknowledge the indispensable help of Dr. Wolfgang *Glänzel* in data processing and building of indicators.

References

1. M. J. MORAVCSIK, How to correct one-dimensional thinking. Appendix A, in: *How to Grow Science*, Universe Books, New York, 1980.
2. M. J. MORAVCSIK, Life in a multidimensional world – Special Report, *Scientometrics*, 6 (1984) 75–86.
3. Comments to the Special Report (Ref. 2) by M.D. GORDON, S.D. HAITUN, J.D. FRAME, V.V. NALIMOV, H. INHABER, I.N. SENGUPTA, C.D. HURT, R. TODOROV, *Scientometrics*, 6 (1984).
4. E. GARFIELD, H. SMALL, Michael J. Moravcsik: Multidimensional scholar and hero of the Third World Science, *Scientometrics*, 20 (1991) 19–24.
5. A. SCHUBERT, T. BRAUN, Three scientometric etudes on developing countries as a tribute to Michael Moravcsik, *Scientometrics*, 23 (1992) 3–19.
6. T. BRAUN, W. GLÄNZEL, A. SCHUBERT, Scientometric Indicators Datafiles, *Scientometrics*, 28 (1993) 137–150.
7. A. SCHUBERT, W. GLÄNZEL, T, BRAUN, Scientometric Datafiles. A comprehensive set of indicators on 2649 journals and 96 countries in all major science fields and subfields, *Scientometrics*, 16 (1993) 3–478.
8. T. BRAUN, W. GLÄNZEL, H. MACZELKA, A. SCHUBERT, World science in the eighties. National performance in publication output and citation impact, 1985–1989 versus 1980–84. Part I. All science fields combined, physics and chemistry, *Scientometrics*, 29 (1994) 299–334.
9. T. BRAUN, W. GLÄNZEL, H. MACZELKA, A. SCHUBERT, World science in the eighties. National performance in publication output and citation impact, 1985–1989 versus 1980–84. Part II. Life sciences, engineering and mathematics, *Scientometrics*, 31 (1994) 3–30.
10. M. BONITZ, E. BRUCKNER, A. SCHARNHORST, Measuring the Matthew effect. (to be publishded)
11. T. BRAUN, A. SCHUBERT, The landscape of national performances in the sciences, 1981–1985, *Scientometrics*, 20 (1991) 9–17.
12. J. BEHRENS, Y. C. HO, Visualization techniques of different dimensions, found at URL http://seamonkey.asu.edu/~behrens.

Appendix 1
ISO standard country codes of 44 countries

ARG	Argentina	JPN	Japan
AUS	Australia	KOR	South Korea
AUT	Austria	MEX	Mexico
BEL	Belgium	NGA	Nigeria
BGR	Bulgaria	NLD	Netherlands
BRA	Brazil	NOR	Norway
CAN	Canada	NZL	New Zealand
CHE	Switzerland	POL	Poland
CSK	Czechoslovakia	PRC	PR China
DEU	Germany FR	PRT	Portugal
DNK	Denmark	ROM	Romania
EGY	Egypt	SAU	Saudi Arabia
ESP	Spain	SGP	Singapore
FIN	Finland	SUN	USSR
FRA	France	SWE	Sweden
GRC	Greece	TUR	Turkey
HKG	Hong Kong	TWN	Taiwan
HUN	Hungary	UKD	UK
IND	India	USA	USA
IRL	Ireland	VEN	Venezuela
ISR	Israel	YUG	Yugoslavia
ITA	Italy	ZAF	South African R