ANIMAL BEHAVIOUR, 2005, 69, 1399–1413
doi:10.1016/j.anbehav.2004.08.020

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Trends in animal behaviour research (1968–2002): ethoinformatics and the mining of library databases

TERRY J. ORD*, EMÍLIA P. MARTINS*, SIDHARTH THAKUR†, KETAN K. MANE‡ & KATY BÖRNER‡
*Department of Biology and Centre for the Integrative Study of Animal Behavior, Indiana University
†Computer Science Department, Indiana University
‡School of Library and Information Science, Indiana University

*(Received 8 March 2004; initial acceptance 5 April 2004;
final acceptance 31 August 2004; published online 7 April 2005; MS. number: A9835R)*

We applied modern bioinformatic tools to titles and keywords of animal behaviour publications contained in an electronic library database to examine trends in animal behaviour research. We provide the first quantitative overview of animal behaviour research covering 42 836 documents published in the last three decades, across 25 journals. Our study confirms several patterns noted by previous reviews, and offers several novel insights into the history of our field. Profound historical distinctions between early ethology and comparative psychology have been recently bridged by shared interest in communication and social behaviour, and research from physiology and applied areas. Although we reiterate the rise of sexual selection and mating behaviour as prominent areas of research, we also show that interest in mechanism and development has proven particularly resilient over the years. Currently, researchers at hundreds of institutions worldwide are studying animal behaviour. Domesticated animals, foraging/dispersal and learning/memory are topics that appear most frequently in publications from regions that have little history of animal behaviour research, suggesting that these subjects are central to the early development of the discipline. Overall, the study of animal behaviour is healthy, growing, and becoming progressively more integrative over time.

The study of animal behaviour has a long and illustrious history (Durant 1986; Dewsbury 1989) featuring many of the great names in science (e.g. Darwin 1872; Tinbergen 1951; Thorpe 1956; Lorenz 1965). Our discipline has undergone remarkable change, particularly over the last few decades. Recent reviews note the rise in popularity of behavioural ecology (Alcock 2003; Bateson 2003), and debate the potential consequences of this apparent skew on other areas of animal behaviour research (e.g. the study of mechanism and development: Bateson & Klopfer 1989; Stamps 2003; West et al. 2003). There are many practical and social factors that may influence the types of questions addressed by behavioural researchers (e.g. modern advances in life sciences technology, global information exchange over the internet). Until recently, however, these factors were exceedingly difficult to consider in any comprehensive way. Here, we make use of modern informatics techniques and electronic library databases to study the recent history of published animal behaviour literature. We first validate our approach by documenting patterns noted by previous reviewers (e.g. the rise of behavioural ecology), and then apply our methods to explore the relative importance of disciplinary and geographical factors.

In the late 19th century, the study of animal behaviour was largely restricted to amateur naturalists (Durant 1986). The discipline emerged formally with the appearance of *Zeitschrift für Tierpsychologie* (1937; now *Ethology*), *Behaviour* (1948) and the *British Journal of Animal Behaviour* (1953; now *Animal Behaviour*). By 1973 the Nobel Prize committee had recognized the pioneering work of Niko Tinbergen, Konrad Lorenz and Carl von Frisch (1953, 1967), and the study of animal behaviour had become well established. We now have sophisticated methods for examining the genetic and physiological underpinnings of behaviour (e.g. Landgraf et al. 2003; Phelps & Young 2003; Cooper & Goller 2004). Technological advances have facilitated the study of function (e.g. acoustic and video playbacks: McGregor 1992; Ord & Evans 2002) and evolution (e.g. the comparative method: Gittleman 1989; Martins 1996), as well as improving techniques for

Correspondence and present address: T. J. Ord, Division of Biological Sciences, Section of Evolution and Ecology, University of California, Davis, CA 95616, U.S.A. (email: tjord@ucdavis.edu).

0003–3472/04/$30.00/0

studying behaviour in the field (Altmann & Altmann 2003). Mathematical and computer modelling have become invaluable tools for predicting how animals will interact with other individuals (e.g. game theory: Maynard Smith & Price 1973) and the environment (e.g. foraging theory: Stephens & Krebs 1986).

The 50-year anniversary of *Animal Behaviour* prompted several recent reflections on the current status and future of the field (Alcock 2003; Altmann & Altmann 2003; Bateson 2003; Slater 2003; Stamps 2003; West et al. 2003). These reviews note the prominence of behavioural ecology (Alcock 2003; Bateson 2003), but also a renaissance of studies on mechanism and development (Stamps 2003; West et al. 2003). However, these commentaries are necessarily based on subjective overviews of only a small proportion of the published literature. Even limiting a survey to only those studies appearing in the last three decades, there are literally hundreds of academic journals that have collectively published tens of thousands of papers (e.g. Table 1). Previous reviews therefore rely on a small subset of material (e.g. Dewsbury 1992; Klopfer & Polemics 2002; Alcock 2003; Huntingford 2003; Salvador et al. 2003) and/or personal reflections (e.g. Bateson & Klopfer 1989; Bateson 2003). In this study, we take a different approach, using specialized computer software to mine library information databases for important historical and contemporary trends. Innovations in information science have spurred the development of a plethora of software to exploit the huge quantities of information offered by new online databases. Although these methods are necessarily limited by the availability of electronic information, they offer scholars a potentially more objective view of the available literature, and a comprehensive way to explore the current focus and historical content of research.

There are numerous articles providing excellent discussion on the potential social and political issues underlying apparent direction shifts in animal behaviour research (Durant 1986; Sherman 1988; Barlow 1989; Dawkins 1989; Dewsbury 1989, 1992, 1994; Armstrong 1991; Alcock & Sherman 1994; Alcock 2003; Bateson 2003; Slater 2003; Stamps 2003; West et al. 2003). To begin our

Table 1. Number of article records available for each journal in this study

	Start of coverage	Incomplete or no coverage	Total documents	1971–1972	1981–1982	1991–1992	2001–2002
Core ABA journals							
Animal Behaviour	1968		5887	154	226	382	491
Behavioral Ecology and Sociobiology	1976	1977–1982, 1985–1991	2011			208	262
Applied Animal Behaviour Science	1986		1700			174	235
Behavioral Ecology	1990		928			81	206
Behaviour	1968		2000	61	81	131	160
Ethology	1991		870			97	158
Behavioural Processes	1976	1977	1244		52	103	131
Journal of Insect Behavior	1988		850			104	118
Animal Learning & Behavior	1973	1978–1980	1500		146	96	64
Journal of Experimental Psychology: Animal Behavior Processes	1975		780		24	74	64
Learning & Motivation	1971		888	52	53	37	53
Journal of Ethology	1986		302			36	40
Bird Behavior	1981	1982–1983, 1988, 1992–1995	112		4	16	8
		Subtotal	19 072	267	586	1539	1990
Supplement journals							
Hormones and Behavior	1971		1496	53	69	83	564
Physiology & Behavior	1968		11 120		639	744	539
Behavioural Brain Research	1980		3257		137	239	502
Behavioral Neuroscience	1984		1931			195	228
Journal of Comparative Psychology	1984	1986–1990	628			88	102
Behavior Genetics	1970	1970–1972, 1977	1178		76	208	86
Learning & Memory	1998		219				85
Aggressive Behavior	1974	1975–1978	652		22	55	70
Journal of the Experimental Analysis of Behavior	1968		2186	178	126	120	70
Ethology Ecology & Evolution	1989		433			68	59
Behavior Research Methods	1984		524			48	50
Behavioral and Brain Sciences	1981	1981–1991, 2002	140		2	15	24
		Subtotal	23 764	231	1071	1863	2379
		Total	42 836	498	1657	3402	4369

'Core ABA journals' are those chosen by Animal Behavior Abstracts (CSA, Inc. 2002), not including The *Birds of North America* or *Etologia*. 'Supplement journals' refer to additional journals included in our analyses despite not being listed as core serials by ABA list (see text for details).

own study, we first confirm the validity of our approach by documenting some of the patterns shown by earlier authors (e.g. distinctions between classical ethology and comparative psychology (Dewsbury 1989, 1992), and the expansion of behavioural ecology (Bateson & Klopfer 1989; Dawkins 1989; Alcock 2003; Bateson 2003)). We then consider several entirely new questions, for example, measuring the extent to which initial differences between ethology and comparative psychology have been bridged, questioning the decline and/or resurgence of 'mechanism' studies, and exploring the role of specific journals in changing our field. We also offer a contemporary snapshot of our discipline in terms of the research being done, where it is being published, and who is producing it.

To conduct our quantitative overview of animal behaviour research, we applied data-mining and knowledge visualization tools to citation records available online in the Biological Abstracts database (BioSciences Information Service, BIOSIS, Inc. Philadelphia, Pennsylvania, U.S.A.). Of the 25 journals that exhibited a primary focus on the study of behaviour, we found 42 836 records for documents published from 1968 to 2002. First, we adopted a historical perspective and mapped major trends in research interest over the last three decades. We used similarities in title vocabulary to identify research topics that linked individual studies into clusters and explored how these clusters have changed over time. We also conducted a 'burst' analysis to identify dramatic increases and/or sustained usage of particular terms and to track changes in popular title-words over time. Second, we shifted our focus to keywords and explored patterns in contemporary behavioural research (2001–2002) explained by journal, geographical and institutional factors.

METHODS

Data Collection and Extraction

To identify journals covering animal behaviour research, we began with the serials source list used by the Animal Behavior Abstracts (ABA) database (Cambridge Scientific Abstracts, CSA, Inc., Bethesda, Maryland, U.S.A. 2002). This list covers 235 journals categorized by relevance, including 15 'core' journals in which all material appearing in the journal is pertinent to animal behaviour (e.g. *Animal Behaviour*). Focusing on these core journals, we removed two: there were no publication records available for *Etologia*, and *The Birds of North America* (a natural/life history reference) seemed too general in scope to be included in our analyses. There were several conspicuous absences from the ABA core list, most notably the *Journal of Comparative Psychology* (not appearing anywhere on the ABA list) and physiology-oriented journals, such as *Hormones and Behavior* (usually listed by ABA as 'priority', reflecting that these serials sometimes publish research that could not be strictly considered 'animal behaviour'). We therefore supplemented the reduced ABA list with 12 additional journals: the *Journal of Comparative Psychology* and those that include 'behaviour', 'ethology' or 'learning' in the journal title (Table 1). Although we conducted separate analyses on the 13 core ABA journals and the complete listing of 25 journals, we report results only for the latter unless otherwise specified.

There are many biological journals that publish animal behaviour research not included in our study (e.g. *Nature*, *Science*, *Proceedings of the Royal Society of London, Series B*). However, such journals also produce a massive number of nonbehaviour-related publications, which would bias the results of our analyses and compromise the objectives of our study. Although papers appearing in high-impact journals such as *Nature* or *Science* are important contributions to our field, the value of including these articles in our analyses is negated by technological limitations of having to also incorporate the entirety of the journal's contents. Doing so would obscure most of the interesting patterns related to animal behaviour research and expand the focus of our study from trends in behavioural research to those more relevant to the life sciences as a whole (e.g. see Mane & Börner 2004). The *Proceedings of the National Academy of Sciences* alone published more research articles in the last 20 years of our sample than did the 25 journals included in our study combined (Boyack 2004). A preliminary run of our analyses that included The *Birds of North America* identified 'geographic distribution' as the most common keyword in animal behaviour publications, simply because the large number of publications in this serial, and its consistent use of a particular keyword, swamped patterns from other animal behaviour journals. We struck a balance by supplementing our 'core' animal behaviour journals with ABA 'priority' serials that clearly emphasize behavioural research, despite publishing also more general scientific research. A comparison between these two types of journals gives us some insight into differences between publications in core and peripheral journals, leaving for future studies the task of developing reliable methods for identifying 'behaviour' papers in more general scientific journals (e.g. Griffiths & Steyvers 2004).

Using International Standard Serial Numbers (ISSN) to identify journals, we downloaded all available citation records as ASCII text files from the Biological Abstracts database (BIOSIS, Inc. 2002). The standard library web interface prevents data streaming by restricting the number of records that can be downloaded at a time, which can result in a large number of network time-outs and failures. Instead, we used access client software SilverPlatter WinSPIRS v4.01 (Ovid Technologies, Inc., New York, New York, U.S.A.) to download records directly from the server of Indiana University's main library.

We wrote several parsers to extract relevant data from downloaded files. Parsers are algorithms that identify and break data into smaller elements according to their structure or syntax within a data file. For example, a parser was written to use commas to break up an address string (e.g. T. J. Ord, Department of Biology, Indiana University, Bloomington, Indiana 47405, U.S.A.) into separate sections identifying author, department, university, city, state and country, allowing the extraction of information associated with a specific field (e.g. country). Other parsers resemble the 'find and replace' option available in word-processing software. All parser code and software used in this study are available through the Information Visualization Software Repository (Börner & Zhou 2001; Börner 2004). Using parsers, we converted downloaded files to delimited text files and removed duplicate documents (identified by the string 'title#year#author#keywords'). To verify that our data set was complete, we noted the number of documents listed by Biological Abstracts during each ISSN search and later crosschecked this number with those files actually downloaded. Finally, we calculated summary information on the number of unique records for each journal in each year. Publication records for several journals were incomplete, ranging from approximately 1 year (e.g. *Behavioural Processes*: 1977) to 13 years (e.g. *Behavioral Ecology and Sociobiology*: 1977–1982, 1985–1991; Table 1). Before proceeding with our analyses, we confirmed these records were missing from Biological Abstracts (E. Ten Have, BIOSIS, personal communication) and not the consequence of errors accumulated during data acquisition. Our final data set covered 25 journals from 1968 to 2002 totalling 42 836 records of published material (Table 1, Fig. 1).

Our analysis focuses on the vocabulary chosen for titles and keywords of animal behaviour publications. Keywords and title vocabulary in the animal behaviour literature provide complementary rather than duplicate information. Titles tend to focus on the primary subject of a paper and are more useful in identifying broad themes in animal behaviour research. We used title-words to explore major changes in topics and study organisms of interest across the three decades included in our data set. Keywords contain more detailed information and have the advantage of retaining compound terms, such as 'sexual selection' and 'parental care'. We used keywords for more detailed analysis of journal and geographical trends.

In our analysis of title-words, we began by using parsers to remove uninformative words such as 'of' and 'the', identified from stop-word lists (available from Börner & Zhou 2001; Börner 2004). Keyword information is presented in Biological Abstracts as a string of compound words separated by semicolons (e.g. 'anthophilous insects; breeding systems; climate severity; disturbance; evolution; habitat') and are found in two separate fields: DE or 'descriptors', and MI or 'misc. indicators'. We combined keywords from DE and MI fields before conducting further analysis. For both title and keyword analyses, we also removed 'behavio/ur' and 'behavio/ural', which occurred often enough that they might obscure more subtle document associations.

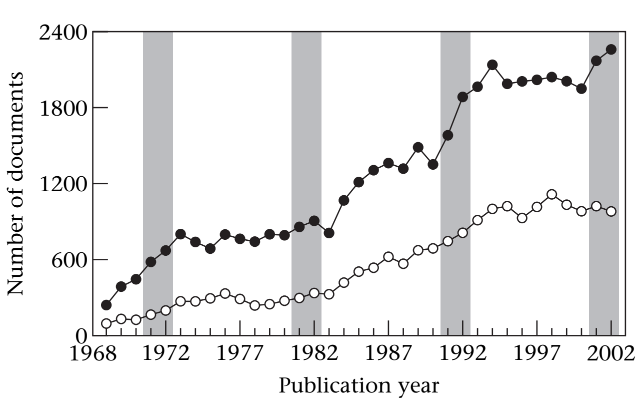


Figure 1. The number of documents covered in the Biological Abstracts database that appear in journals publishing animal behaviour research. Open circles indicate document counts for journals considered to be core contributors to Animal Behavior Abstracts, while closed circles represent counts for all journals included in this review (see Table 1). Shaded areas represent 2-year 'snapshots' used in some document analyses.

Knowledge Domain Visualizations

Relationships among prominent topic areas through time

To identify broad themes in animal behaviour, and to monitor their changes over time, we entered title-words into a latent semantic analysis (LSA; Landauer et al. 1998). LSA produces a large matrix of documents by title-words and applies a singular value decomposition (using LSA SVDPACKC; Berry 1993) to highlight important latent dimensions in the matrix. Put simply, the analysis isolates each word in a title and searches for the same term in other document titles. When the program finds common vocabulary, it creates a link between the documents. The more links shared between records, the greater the assumed overlap in topic area (see below).

There were 24 850 unique title-words in our data set, a sample size too large to be effectively manipulated by visualization software. To explore changes in prominent research themes and topic relationships through time, we isolated publications appearing in 1971–1972, 1981–1982, 1991–1992 and 2001–2002 for all 25 journals (Fig. 1, Table 1), conducting separate analyses for each pair of years. These 'snapshots' reduced the data set to a manageable size, while still providing a broad sample to illustrate research trends reliably.

In essence, LSA produces a similarity score for each comparison of two publication titles based on the number and context of common words. The strength of LSA lies in its ability to resolve the potentially confounding problem of synonymy (similar meaning words) and polysemy (words with multiple meaning) by comparing, not only the similarity in vocabulary, but also the context in which words appear. Hence, document titles with common words used in similar contexts will be judged to be more closely related than if they simply share identical terms. Text similarity judgements obtained using LSA are consistent with human word sorting and category judgements (Landauer et al. 1998).

We used Pajek visualization software (Batagelj & Mrvar 1997, 2003) to identify clusters of related documents and links among them based on the most important latent dimensions. Implementing the 'Kamada–Kawai' algorithm (Kamada & Kawai 1989) with circular start position, this program was used to draw two-dimensional images of the document similarity matrices, effectively highlighting groups or clusters of documents with similar title vocabulary. The program begins by placing documents (represented by a node or dot) along an outer circle. It then sifts through the similarity matrix produced by the LSA and connects publications with a line according to their similarity values. As links are created, documents are drawn together. By providing a visual representation of matrices (hence 'visualization'), we can easily identify how documents are related to each other.

To reduce clutter in plots caused by superfluous links between documents, only similarity scores greater than 70% were traced. The success of particular thresholds to draw out salient document associations is dependent on the number of documents analysed: the larger the pool of samples, the greater the chance an individual record has of sharing semantic similarity with another document. Because the number of articles published in each 2-year snapshot varied (Table 1), we tailored similarity cutoffs for each cohort to optimize the visibility of clusters. Final cutoffs were 0.70, 0.70, 0.74 and 0.755 for 1971–1972, 1981–1982, 1991–1992 and 2001–2002, respectively. Last, we labelled each observed cluster by manually isolating individual documents that occurred unambiguously within a cluster, extracting the title-words for these documents and identifying which shared words were the most common (and hence predominantly responsible for the aggregation). While the overall two-dimensional projection of documents holds no inherent meaning (e.g. the circular start position of nodes at the beginning of the plot-rendering process), the spatial layout of clusters relative to each other does allow the identification of links between document groups and a way to quantify relationships between prominent research areas.

Emergence and longevity of new topic areas

Next we applied a 'burst' detection algorithm (Kleinberg 2002) to explore how major themes in animal behaviour research have changed across time. Kleinberg's (2002) model was originally designed to sort email automatically into meaningful folders by identifying important topics as they appear, grow in popularity for a period of time and then fade away. As applied here, the burst detection algorithm focuses on the temporal intervals between repeated appearances of the same term. When a term is popular, it will be used frequently and the time intervals between repeated appearances will be short. The two-state form of the burst detection algorithm finds the model that best describes the data as a collection of temporal strings of high (i.e. bursts) and low episodes of popularity for each of the terms studied. 'Weights' are also calculated to allow for direct comparison among bursts for the same and different words in terms of their relative prominence.

Returning to the complete set of 42 836 publications appearing over 35 years, we excluded words such as 'the' and 'an' to obtain a total of 24 850 unique title-words. We further focused our attention on the 739 title-words that appeared at least 100 times in the data set. To these, we applied the burst detection algorithm, looking across the full 35 years of publications, to identify rapid increases and decreases in popularity (bursts) for each term through time.

Journal, geographical area, and institutional patterns

For more detailed consideration of current journal, geographical and institution coverage, we shifted our focus to keywords, considering only terms reported for the most recent publications (2001–2002; 4369 records). First, to assess journal coverage, we identified and compared the most common keywords used by each of the 25 journals of interest. Second, for examining differences in research trends across the world, we used the institutional address of the first author to identify the country of origin. We excluded all records with ambiguous affiliations resulting from missing words or country abbreviations and grouped the subsequent 57 identifiable countries into eight geographical regions to facilitate interpretation. Finally, the five-digit zip code of the U.S. is also unusually informative as it allows the identification of specific institutions (each institution is typically given a unique zip code). Using the address of the first author, we extracted documents published by scholars in the U.S. Zip codes were then examined to reveal those institutions that produced the most animal behaviour publications over the 2001 and 2002 period. Unfortunately, we unable to conduct a similar analysis on a global scale, in part, as zip (or post) codes in many countries are less likely to be specific to particular institutions. Institutional names proved even more difficult to isolate reliably because of variation in how the names were abbreviated and where they were placed within address strings.

RESULTS

The number of documents published in 25 animal behaviour journals has increased steadily from 1968 to 2002 (229 to 2254 documents in each year, respectively; Fig. 1). *Animal Behaviour* is by far the largest recent contributor of the 13 core journals covered by Animal Behavior Abstracts (11.2%; Table 1), followed by *Behavioral Ecology and Sociobiology*, *Applied Animal Behaviour Science*, *Behavioral Neuroscience* and *Behavioral Ecology*, which contribute 5–6% each. Note that roughly half of the publications included in our study appeared in journals not listed by ABA as 'core' sources, irrespective of time frame considered (Table 1).

Topic Area Associations Identified by Title-words

The vocabulary used in publication titles has become dramatically diverse over the last three decades, in part because of a huge increase in the number of publications. The occurrence of unique title-words in our data set increased from 1540 title-words in the 1971–1972 snapshot to 2238 title-words in the 2001–2002 snapshot. As a consequence, identifiable clusters in our visualizations also increased from approximately nine in 1971–1972 to over 20 in 2001–2002 (Fig. 2).

The language relationships appearing in publication titles for 1971–1972 and 1981–1982 reflects a profound separation between early ethology and comparative psychology. In 1971–1972, there is one cluster focused on function (e.g. 'feeding', 'sexual', 'aggression': cluster 1), two clusters concentrating on animal learning (e.g. 'stimulation', 'schedule', 'reinforcement', 'conditioned':

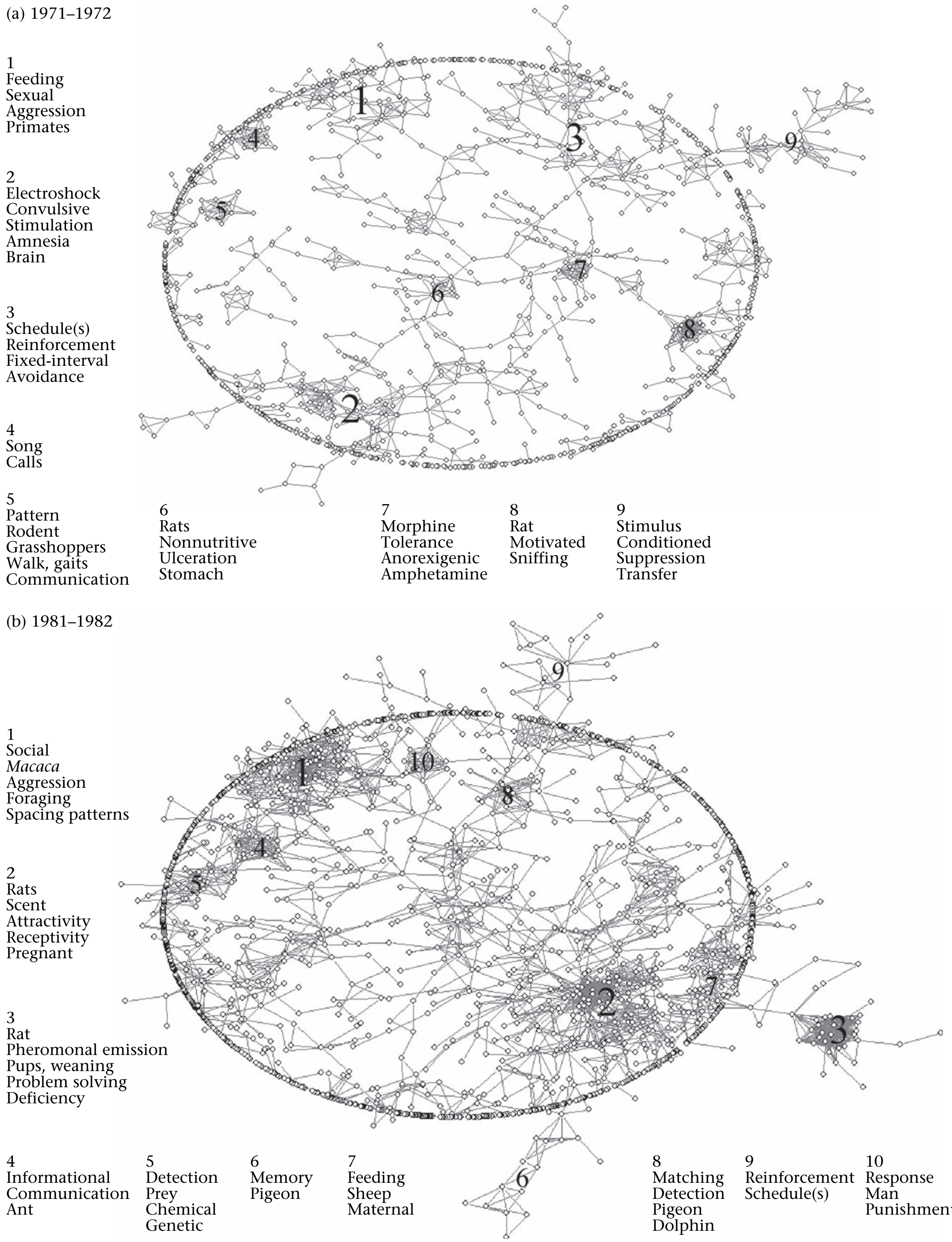


Figure 2. Visualization plots of latent semantic analyses for documents published in all 25 behaviour journals for (a) 1971–1972, (b) 1981–1982, (c) 1991–1992 and (d) 2001–2002. Nodes represent individual papers, with connecting lines indicating similarity in title vocabulary. Common areas of research form clusters. Salient terminology producing prominent clusters are given. Those nodes left on the periphery are generally unrelated to records within the document space. See text for details.

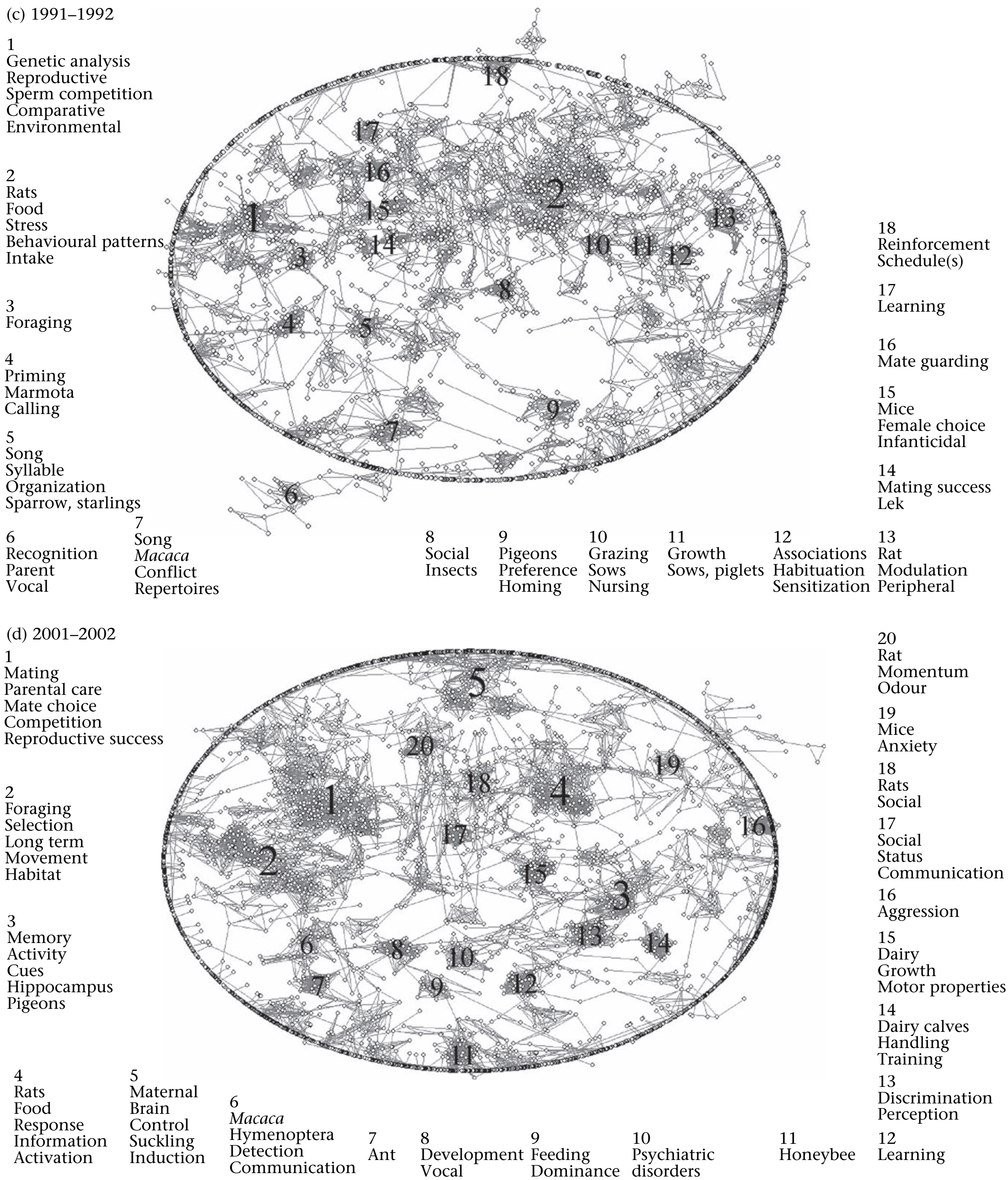


Figure 2. (*continued*)

clusters 3 and 9), and several other groups centred on specific experimental techniques (e.g. 'electroshock', 'ulceration', 'amphetamine'; Fig. 2a). In 1981–1982 (Fig. 2b), there was prominent clustering around 'social', 'aggression' and 'foraging' (cluster 1), with smaller connected clusters featuring 'communication' and 'prey'. Interestingly, 'genetic' and 'chemical' also appeared to be associated with these groups. Animal learning is now loosely united across clusters 8 ('matching', 'detection') and 9 ('reinforcement'), and comparative psychologists have identified a preferred model organism with two large clusters (clusters 2 and 3) joining documents describing research on ‘rat/s’. Both animal learning and ‘rat/s’ groups are only distantly connected to ethologically oriented publications (cluster 1), emphasizing the continuing distinction between ethology and comparative psychology into the 1980s.

Between 1991 and 1992 there was a considerable increase in heterogeneity of title vocabulary (Fig. 2c), with 18 clearly identifiable clusters reflecting diversification across a multitude of topic areas. There was still a large, linked cluster describing ethological studies of animal behaviour (cluster 1), although with a greater emphasis on genetics and reproduction, in addition to ‘sperm competition’ (a previously unseen topic area). As in earlier years, a second large cluster focused on ‘rats’ (cluster 2). However, these clusters are now linked through several smaller groupings focusing on reproductive behaviour and mate choice (clusters 14–16). Appearing in four clusters, sexual selection has emerged as a fashionable topic area. Acoustic communication has also grown, featuring heavily in several clusters (clusters 4–7). Other clusters indicate a new interest in applied animal behaviour research (e.g. clusters 10 and 11: ‘sows’, ‘growth’ and ‘nursing’).

Continued diversification is evident in 2001–2002. The emphasis on reproductive behaviour and sexual selection seems to have expanded to other areas such as ‘parental care’ (cluster 1). A second large and related cluster describes greater interest in ‘foraging’, ‘movement’ and ‘habitat’ (cluster 2). The large ‘rats’ cluster has now become subdivided into separate groups describing specific types of behaviour, including: ‘memory’ (cluster 3); ‘food’ and ‘response’ (cluster 4); ‘maternal’, ‘brain’, and ‘control’ (cluster 5); and ‘learning’ (cluster 12). Several new groups indicate a concentration of research on specific model and domesticated organisms (e.g. ‘bees’, ‘ants’, ‘mice’, ‘dairy calves’). Interestingly, the remnants of the division between ethology (e.g. clusters 1 and 2) and comparative psychology (e.g. clusters 3 and 4) are now bridged primarily by clusters relating to communication and social behaviour (clusters 17 and 18), and a growing cluster referring to rat and odour (cluster 20). Applied animal behaviour (clusters 14 and 15) also remains well represented and offers new links to the comparative psychology clusters.

Bursts in Popularity of Title-words

Focusing our attention on the 739 title-words that appeared at least 100 times in our data set, the burst detection algorithm identified 506 bursts of popular title-words across 35 years. Bursts were regularly spaced, lasting a median of 4 years (mean $\pm$ SE $= 5.6 \pm 0.19$). There were 470 title-words such as ‘effect’, ‘role’ or ‘difference’ that are difficult to interpret further. If we focus on the remaining 269 terms, there appear to be three vocabulary periods: pre-1985, 1985–1995, and post-1995 (Fig. 3). Terms that burst in popularity during the 1985–1995 ‘transition’ period tended not to burst before or afterwards.

Of the 269 potentially meaningful terms, 200 reflect topics of interest in animal behaviour research (e.g. ‘operant’, ‘evolution’, ‘predation’). The words occurring within each of the three time periods cross disciplinary boundaries, indicating the continuing diversity of animal behaviour research throughout the history of our field. The early periods, for example, show bursts from ‘shock’, ‘reinforcement’, ‘natal’ and ‘testosterone’. ‘Guarding’, ‘genetic’, ‘anxiety’ and ‘opioid’ all burst during the transition time period (1985–1995). ‘Receptor’, ‘anxiety’, ‘paternity’ and ‘mate’ all burst in the most recent time interval.

The 69 remaining terms refer to a type of animal (e.g. ‘rats’ appeared 5550 times; Fig. 3b), and also reflect some meaningful shifts over the years. Before 1985, virtually all animal terms undergoing bursts of popularity were model organisms, including cats, monkeys, squirrels (which could also be ‘squirrel monkeys’) and chickens (Fig. 3b). In the 1985–1995 transition period, there were several bursts referring to insects (especially hymenoptera (e.g. bees, wasps and ants) and orthoptera (e.g. crickets, grasshoppers and katydids)) that were not abundant earlier. In the mid-1990s, there was a sudden surge of interest in a more diverse group of domesticated animals (e.g. dogs) and those of economic importance (e.g. cows, deer).

Journal Focus

Listing the 10 most frequently reported keywords for each of the 25 journals in 2001–2002, we found a total of 143 different terms and only a moderate degree of overlap across journals. Learning, memory, reinforcement and related terms appeared consistently in the top-10 keyword lists of all but a few journals, emphasizing the continuing importance and broad impact of these themes to animal behaviour research. Mate choice, reproductive success and sexual selection also appeared frequently, indicating the popularity of evolutionary questions. When the most common keywords for all journals were pooled and ranked, terms referring to aggression, learning/memory, foraging and sexual selection were at the top of the list (Fig. 4a).

Keywords also showed the existence of a continuum between journals that emphasize evolutionary ethology and those that publish comparative psychology. On one extreme, *Behavioral Ecology* and *Behavioral Ecology and Sociobiology* place an unusually strong emphasis on ‘sexual selection’, and other terms related to reproduction and mating (Fig. 4b). *Animal Behaviour*, *Behaviour* and *Ethology* also publish articles on sexual selection, but add social behaviour, predation, foraging, communication (usually ‘vocalization’) and evolution. At the other extreme, nearly all popular keywords reported by *Behavioural Processes* and *Animal Learning & Behavior* address learning and memory (Fig. 4c).

Interestingly, *Applied Animal Behaviour Science* lies somewhere in between; including ‘aggression’ and ‘vocalization’ as prominent keywords, but also ‘stress’ and ‘motivation’. Most of the journals that appeared on our supplemented list of 25, but not in the core list from ABA, also appeared somewhere in the middle. For example, physiology journals like *Hormones and Behavior* frequently publish studies about ‘sexual behaviour’ and ‘aggression’, in addition to 'stress', 'learning' and 'photoperiod'. The *Journal of Comparative Psychology* showed an unusually high peak with 'animal communication' appearing along side research relating to 'habituation', 'development' and 'imitative learning'. These journals therefore form an important bridge between the academic descendants of early ethologists and comparative psychologists. They also tend to be more specialized than journals at the extremes, leading to a more highly skewed distribution of keywords and/or a larger proportion of unique terms (Fig. 4d). For example, the focus of *Applied Animal Behaviour Science* is on 'animal welfare', whereas *Hormones and Behaviour* publishes more papers described by 'neuro/endocrinology'.

Global and Institutional Trends

In 2001–2002, animal behaviour publications were produced by researchers affiliated with institutions in every region of the world (Fig. 5a), with North America and Western Europe being the primary producers of animal behaviour research. Keywords used by North America, Western Europe and Australia/New Zealand were remarkably similar, reflecting global agreement on popular topic areas in animal behaviour (Fig. 5b). The relatively few contributions from remaining regions (South America, Eastern Europe, Asia and Africa) shared an emphasis on animal learning and domesticated animals, indicating

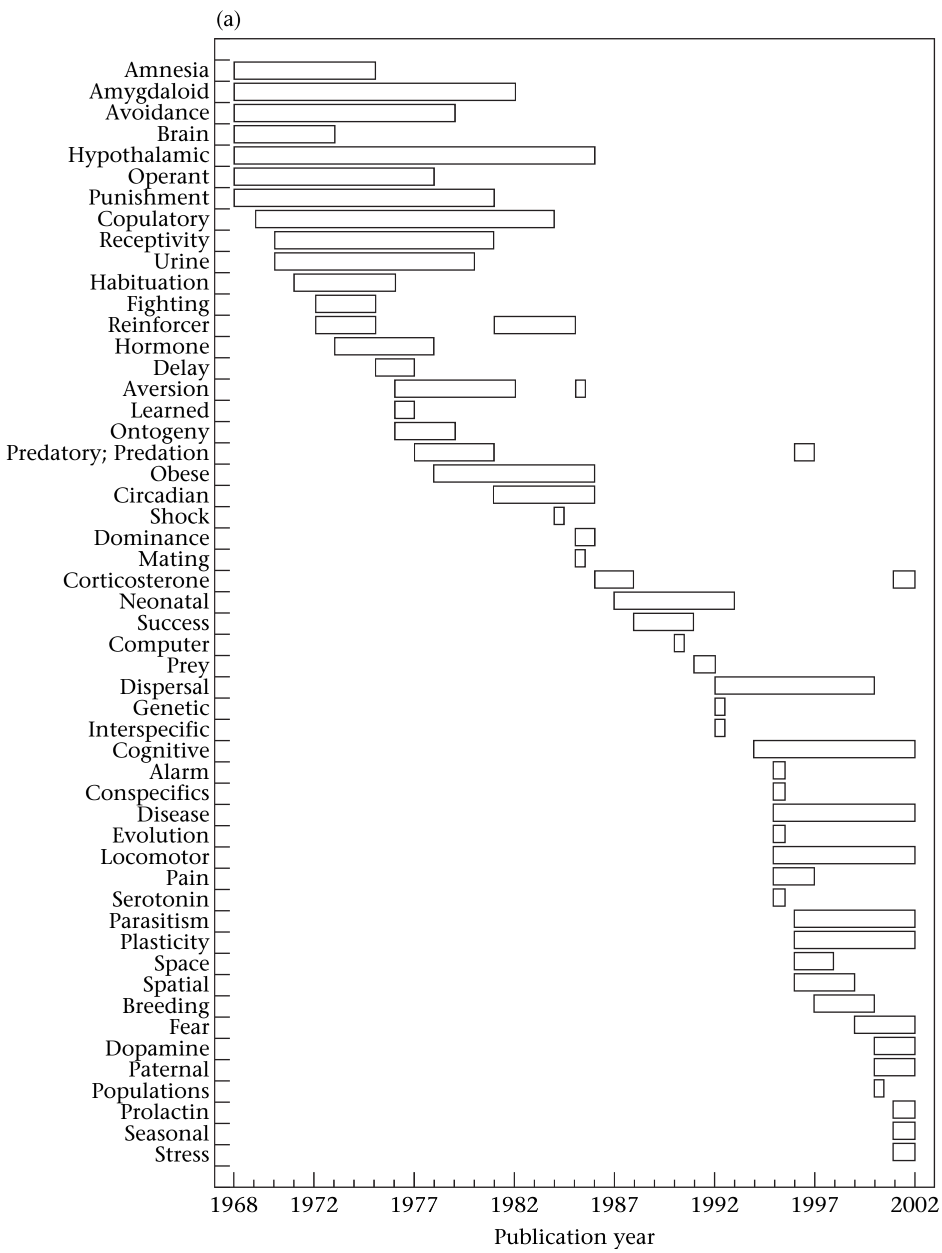


Figure 3. Bursts in popular title-words for documents published in 25 behaviour journals from 1968 to 2002. Vocabulary surges have occurred regularly throughout the last 35 years. (a) However, a subset of the 200 meaningful terms show divisions of vocabulary across three periods: pre-1985, 1985–1995 and post–1995. (b) Popular study organisms also vary over time.

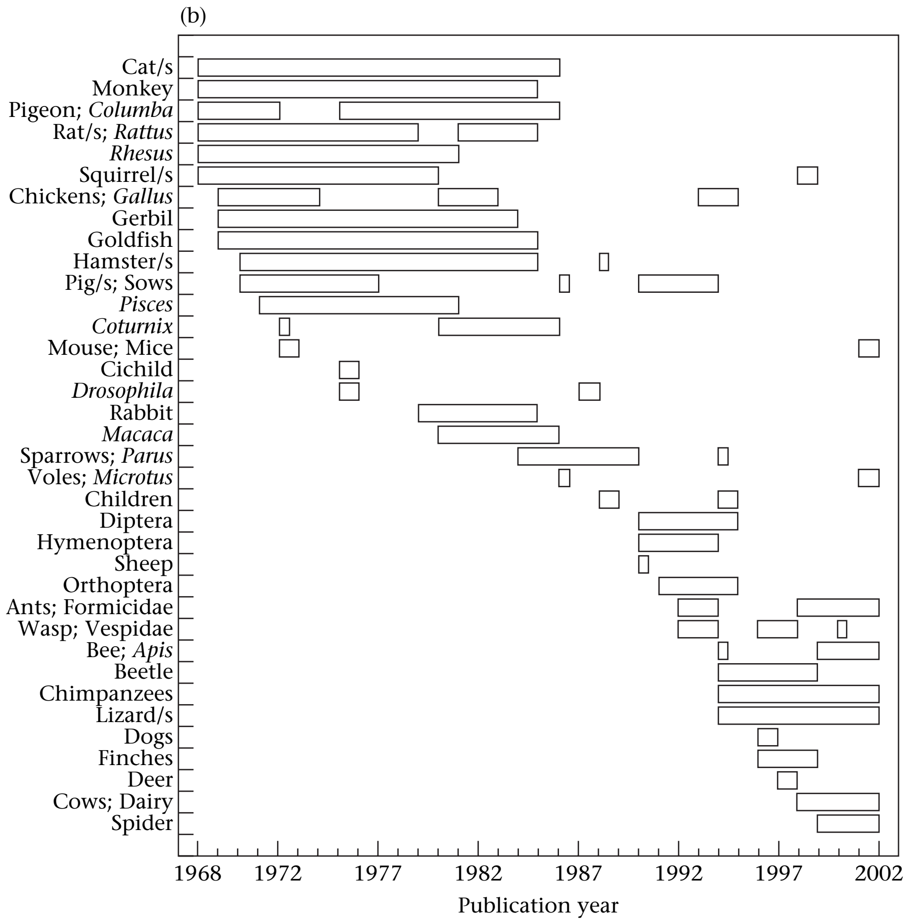


Figure 3. (*continued*)

strong interest in applied animal behaviour research. Relative representation by South America and Asia was larger when considering all 25 journals, whereas Africa, Middle East and Eastern Europe were better represented when only core ABA serials were considered.

In the United States alone, we counted 1806 publications from first authors located at more than 487 different zip codes during the 2001–2002 period. Over 100 zip codes tallied five or more animal behaviour publications, although it is possible that some zip codes do not identify unique institutions (e.g. some could be personal residences). Also, the major representation of some institutions may result from an unusually large number of contributions by single investigators in this 2-year period. The number of documents associated with each zip code also varied between the two lists (13 ABA core journals and complete list of 25). Nevertheless, several institutions known to have larger graduate programmes or a greater number of animal behaviour researchers account for a disproportionate share of publications by geographical location within the U.S. (Table 2).

DISCUSSION

Despite our analyses being necessarily limited by the scope of our study (e.g. the choice of journals, availability of electronic data, interpretations of zip codes), our approach succeeds in documenting many of the patterns mentioned by earlier reviews using very different methods (e.g. Durant 1986; Sherman 1988; Barlow 1989; Dawkins 1989; Dewsbury 1989, 1992, 1994; Armstrong 1991; Alcock & Sherman 1994; Alcock 2003; Bateson 2003; Slater 2003; Stamps 2003; West et al. 2003), while also offering several new insights that have gone previously undetected by traditional review methods. Our findings confirm the historical divisions between biological and psychological branches of animal behaviour research (Dewsbury 1989, 1992). However, the gap between ethology and comparative psychology has diminished over time, with frequent present-day interaction across disciplines occurring through a shared interest in communication and social behaviour. We also document the rise in popularity of behavioural ecology (Bateson & Klopfer 1989; Alcock 2003; Bateson 2003), but find a healthy, continued interest in mechanism, learning and development. Finally, our results offer several tantalizing glimpses into the roles of core and peripheral journals, as well as geographical trends in the study of behaviour.

Tinbergen (1963) described the study of animal behaviour as encompassing four main areas of investigation: mechanism, function, development and evolution/phylogeny. Subsequent authors have grouped these topics into 'proximate' and 'ultimate' questions (mechanism/development and function/evolution, respectively). Commentaries on the development of our field typically herald or lament a shift in interest between adaptive and mechanistic/developmental questions (Wilson 1975; Barlow 1989; Dawkins 1989; Bateson 2003; Stamps 2003; West et al. 2003). However, the relevance of this dichotomy has also been debated (Sherman 1988; Armstrong 1991; Dewsbury 1992, 1994; Alcock & Sherman 1994). Our analysis of the vocabulary used by authors to describe their own research shows that, although the popularity of studies addressing evolutionary function (especially sexual selection) has increased in recent years, interest in mechanistic/developmental research has remained strong throughout the history of animal behaviour research. Instead of one research area dominating the field at the expense of another, a more accurate description is that our field has collectively expanded and become integrated across multiple topic areas. Hence, contemporary research still reflects the diversity and comprehensive approach advocated by Tinbergen (1963) and Huxley (1942).

We found a major difference between 'core' and 'priority' animal behaviour journals (those that publish behavioural research frequently, but also other types of research). The physiology and applied journals included in our study appear to play a pivotal role, not only producing a large number of animal behaviour publications (about half of those reviewed in our study), but offering an important intellectual link between disparate arms of our discipline. In contrast, newer journals that emphasize animal behaviour research appear to be more specialized than their older counterparts, emphasizing

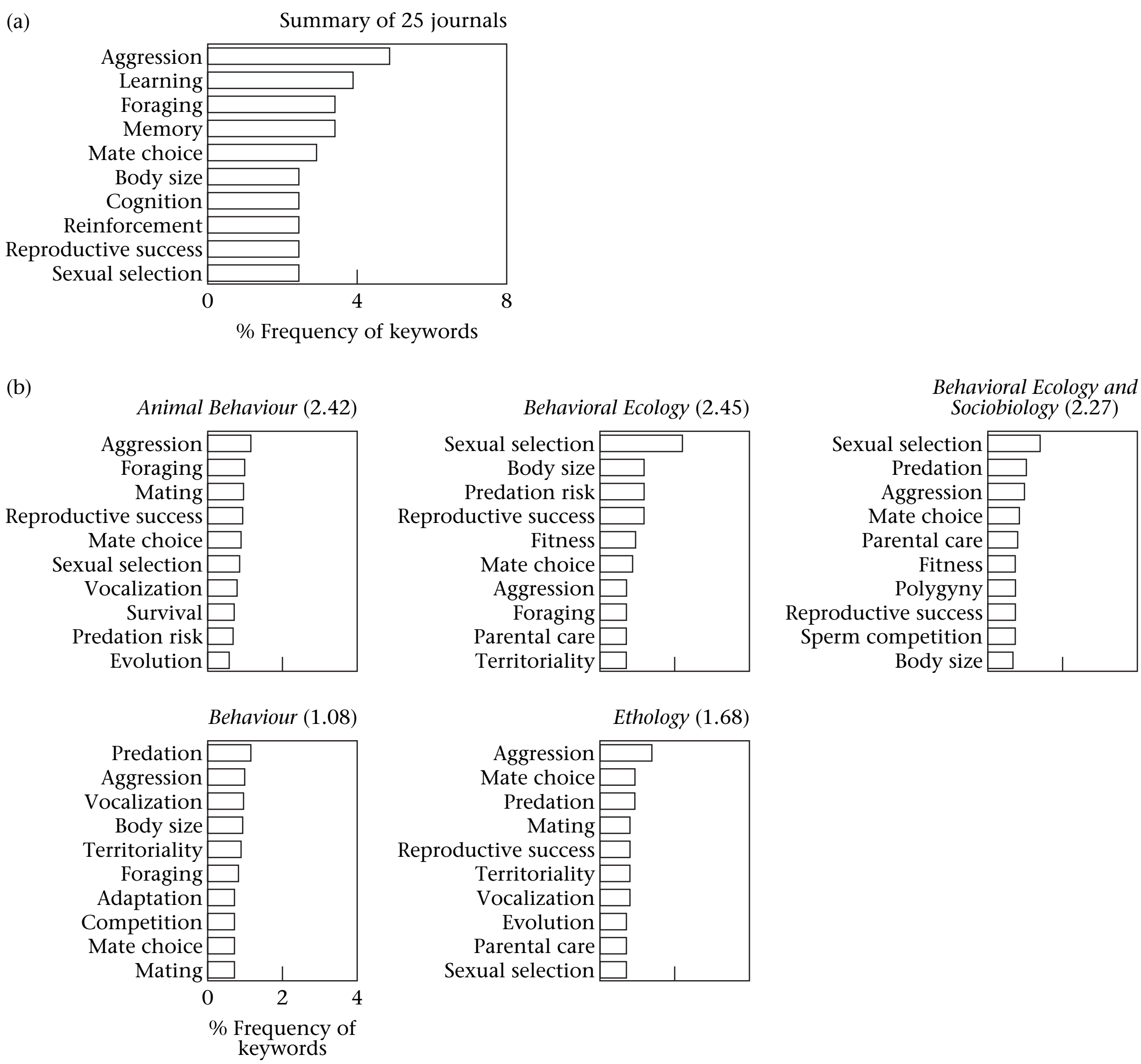


Figure 4. Keyword frequency distribution plots summarizing the content of journals publishing animal behaviour research for 2001–2002: (a) the 10 most frequently used keywords for all journals ranked collectively; (b) journals emphasizing 'evolutionary ethology'; (c) journals largely covering animal learning and memory; and (d) journals that are relatively intermediate in their document coverage. Values in parentheses are journal impact scores calculated by ISI Web of Science for 2002 (Thomson Scientific, Philadelphia, Pennsylvania, U.S.A.; http://www.isinet.com/).

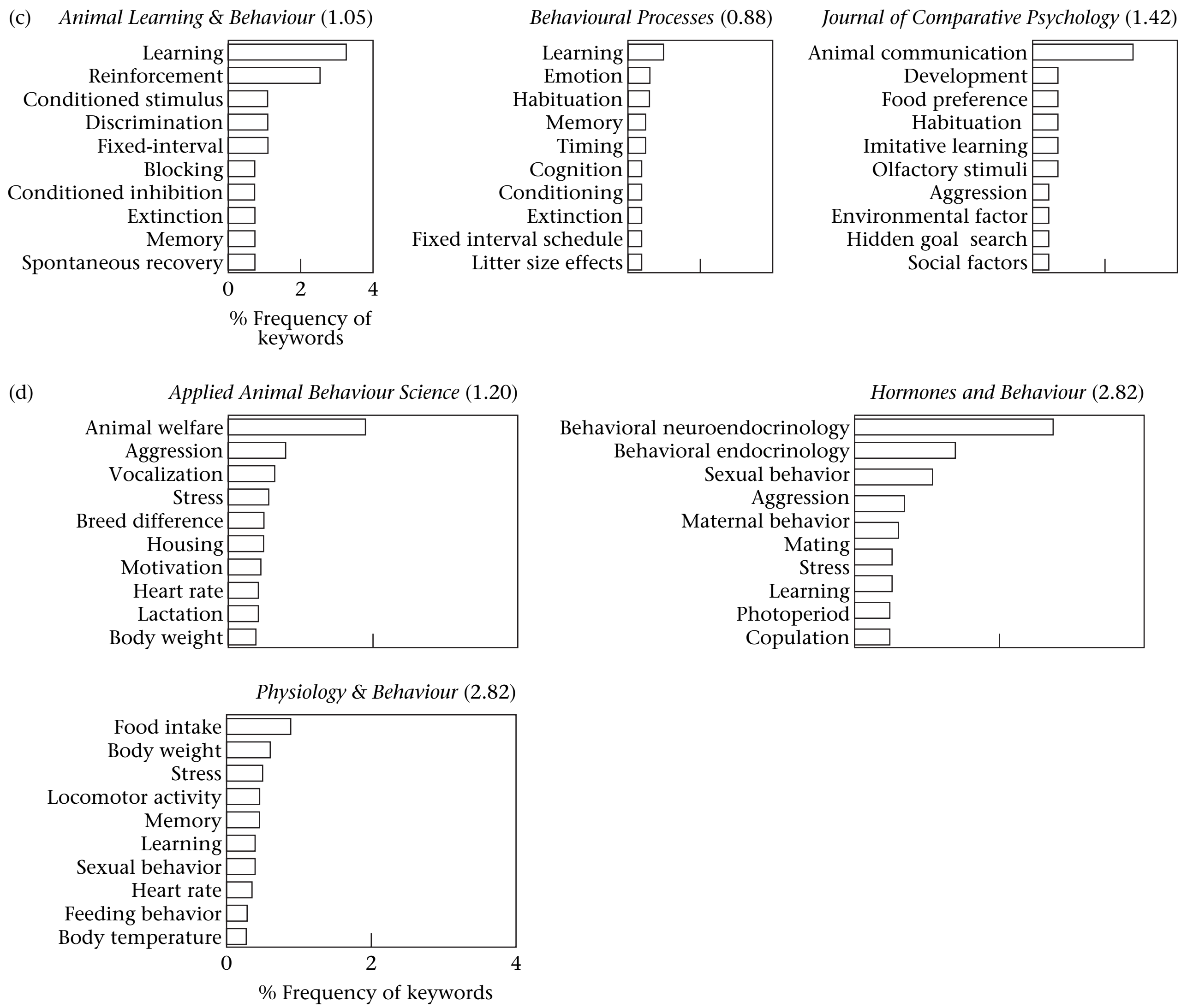


Figure 4. (*continued*)

distinctions between subdisciplines rather than building bridges across them. For example, *Behavioral Ecology* (founded in 1990) and *Behavioral Ecology and Sociobiology* (1976) have a stronger focus on sexual selection and mating than do the older *Ethology* (1937), *Behaviour* (1948) and *Animal Behaviour* (1953). Further analyses exploring the role of the other 200 odd journals that occasionally publish animal behaviour research may provide valuable insight into this phenomenon.

Our results also highlight several intriguing patterns driven by geographical distributions, the impact of particular institutions, and the timing of cultural shifts. We found little evidence for disciplinary differences across geographical regions. North America and Western Europe are very similar in research focus despite the unique historical events in the development of animal behaviour research in these regions (Durant 1986; Dewsbury 1989). Research in countries with little history in the study of behaviour also tend to encompass the breadth of animal behaviour research (i.e. both ethology and comparative psychology), but some geographical regions (Africa, Middle East, Eastern Europe) are better represented in core journals, whereas others (South America, Asia) publish more frequently in journals that do not specialize as much in animal behaviour. Our zip code analysis showed that behavioural research, in the U.S. at least, is widely distributed, with more than 100 institutions contributing five or more animal behaviour publications in a 2-year period. Finally, our results show the rise and fall in popularity of research terms and study organisms in roughly 10-year cycles. As a consequence, the vocabulary and animals examined in the 1970s and 1980s are noticeably different from those studied today.

This study confirms that the study of animal behaviour is active, healthy, and growing. It is being studied by researchers representing hundreds of institutions around the world, including geographical regions with little history of life science research. Sexual selection is clearly an important topic in modern research, but interest in animal learning, aggression, foraging, and other reproductive and

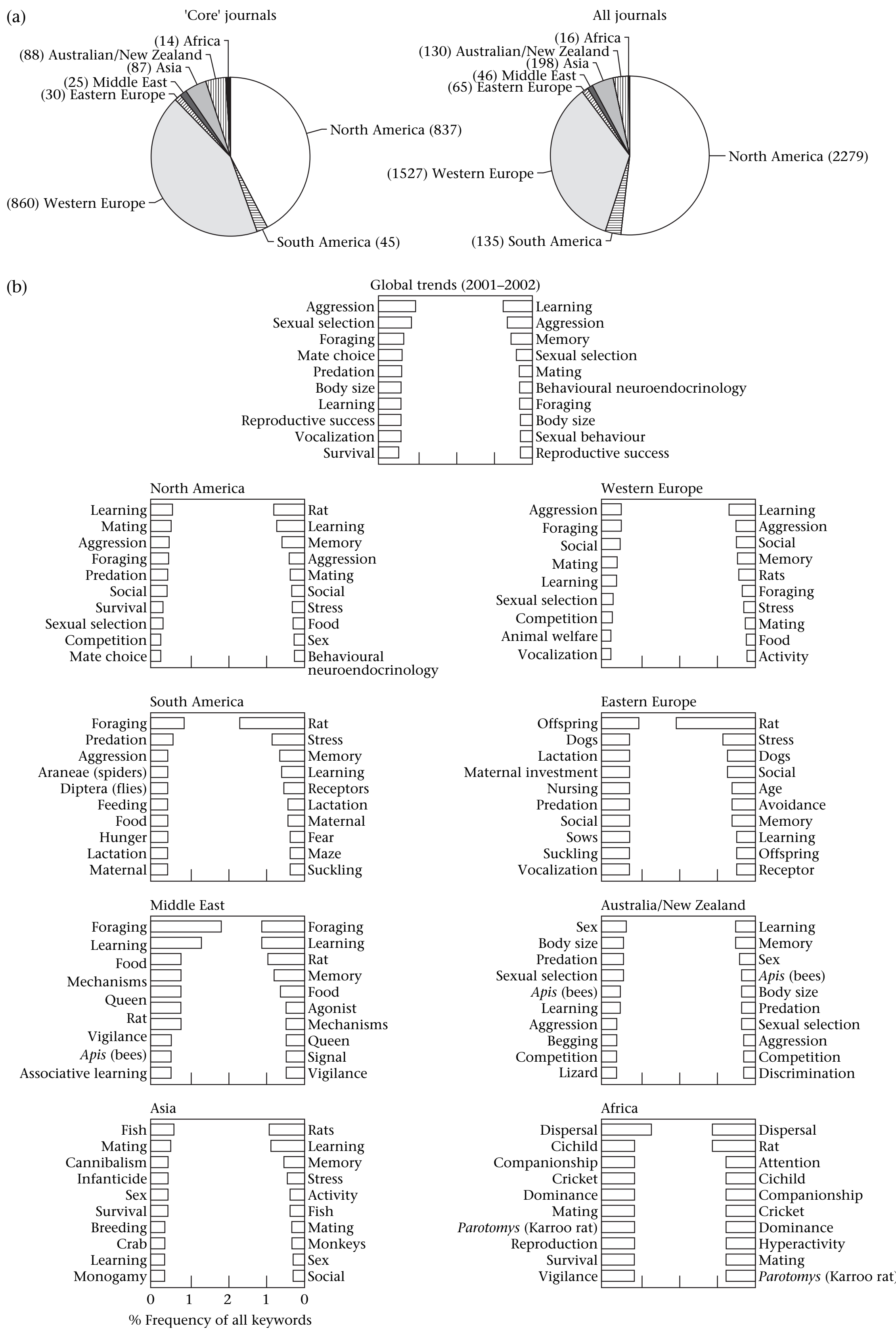


Figure 5. Publication trends in 2001–2002 across the world for both core ABA and supplemented journal listings. (a) The proportion of total documents by region, with values in parentheses representing absolute numbers of documents. (b) Keyword profiles for core (left) and all behaviour journals (right) across the globe and eight separate geographical regions differentiated by principal author address. Histograms are frequency distributions of document keywords.

Table 2. Number of articles published in animal behaviour journals by the first 25 U.S. institutions on our lists for 2001–2002

Core ABA journals (670 total publications)	Document count	All journals (1806 total publications)	Document count
Univ. California, Davis (95616)	20	Univ. California, Davis (95616)	44
Cornell Univ. (14853)	18	Cornell Univ. (14853)	36
Indiana Univ., Bloomington (47405)	14	Indiana Univ., Bloomington (47405)	27
Univ. Memphis (38152)	13	Univ. Michigan (48109)	27
Univ. Kentucky (40506)	13	Univ. Washington (98195)	26
Univ. Washington (98195)	13	Univ. Pennsylvania (19104)	25
Purdue Univ. (47907)	12	Univ. California, Los Angeles (90095)	25
Univ. Michigan (48109)	11	Michigan State Univ. (48824)	24
Univ. Wisconsin-Madison (53706)	11	Purdue Univ. (47907)	23
Arizona State Univ. (85287)	11	Univ. Massachusetts (01003)	22
Univ. California, Los Angeles (90095)	11	Univ. Texas, Austin (78712)	22
Washington State Univ. (99164)	11	Emory Univ. (30322)	21
SUNY, Binghamton (13902)	10	Ohio State Univ. (43210)	21
Brown Univ. (02912)	9	Univ. Pittsburgh (15260)	20
Michigan State Univ. (48824)	9	Florida State Univ. (32306)	20
Univ. Pennsylvania (19104)	8	Univ. Wisconsin-Madison (53706)	20
Univ. Georgia (30602)	8	Arizona State Univ. (85287)	19
Texas A & M Univ. (77843)	8	Boston Univ. (02215)	17
Univ. California, San Diego (92093)	8	SUNY, Binghamton (13902)	16
Univ. California, Berkeley (94720)	8	Univ. Maryland (20742)	16
Univ. Massachusetts (01003)	7	Univ. Georgia (30602)	16
Univ. Maryland (20742)	7	Univ. California, San Diego (92093)	16
Colorado State Univ. (80523)	7	Univ. Kentucky (40506)	15
Harvard Univ. (02138)	6	Univ. California, Berkeley (94720)	15
Princeton Univ. (08544)	6	John Hopkins Univ. (21218)	14

Institutions were identified by the zip code appearing in first author addresses. Counts may include some publications by geographically similar, but unaffiliated researchers (see text for details).

social behaviour remain strong. Animal behaviour research also continues to draw new attention from areas such as domesticated animal science and animal welfare. Moreover, the approach adopted in our study leads to many appealing avenues for future research. Detailed analysis can go beyond titles and keywords to explore the complex web of relationships formed by citation (e.g. White et al. 2004) and collaboration, to investigate how 'schools of thought' develop. Records of awards given by funding bodies could be examined to determine the importance of external grants in directing trends in mainstream research (e.g. Boyack & Börner 2003; Boyack 2004). Similar bioinformatic tools can be applied to organized collections of behavioural measurements, such as those envisioned by EthoSource (a new global initiative to facilitate the sharing and combining of behavioural data; Martins & Clark 2002), and highlight the opportunities for a new field of 'ethoinformatics'.

Acknowledgments

We thank Roger Beckman (Indiana University (IU) Life Sciences Library), Ann Bristow (Reference Department, IU Main Library) and Ralph Quarles (IU Library Information Technology) for providing advice and access to Indiana University's reference databases. Elizabeth Ten Have (Product Manager, BIOSIS) provided additional information regarding Biological Abstracts. We also thank Peter Klopfer for providing us with unpublished material that was the basis of his article in the *ABS Newsletter*, **47(3)**, 4–5; and Barbara Clucas, Britt Heidinger and Erin Kelso for comments on previous versions of the manuscript. We are particularly grateful to George Uetz, Lee Drickamer, Felicity Huntingford and an anonymous referee for discussion that greatly improved this study. This material is based on work supported by National Science Foundation under grant Nos IBN-0205180, IBN-0236049 and DEB-0196357 to E.P.M. and IIS-0238261 and DUE-0333623 to K.B.

References

Alcock, J. 2003. A textbook history of animal behaviour. *Animal Behaviour*, **65**, 3–10.

Alcock, J. & Sherman, P. 1994. The utility of the proximate–ultimate dichotomy in ethology. *Ethology*, **96**, 58–62.

Altmann, S. A. & Altmann, J. 2003. The transformation of behaviour field studies. *Animal Behaviour*, **65**, 413–423.

Armstrong, D. P. 1991. Levels of cause and effect as organizing principles for research in animal behaviour. *Canadian Journal of Zoology*, **69**, 823–829.

Barlow, G. W. 1989. Has sociobiology killed ethology or revitalized it? In: *Perspectives in Ethology, 8* (Ed. by P. P. G. Bateson & P. H. Klopfer), pp. 1–45. New York: Plenum.

Batagelj, V. & Mrvar, A. 1997. Pajek: program package for large network analysis. Available free from: http://vlado.fmf.uni-lj.si/pub/networks/pajek/.

Batagelj, V. & Mrvar, A. 2003. Pajek: analysis and visualization of large networks. In: *Graph Drawing Software* (Ed. by M. Jünger & P. Mutzel), pp. 77–103. Berlin: Springer.

Bateson, P. 2003. The promise of behavioural biology. *Animal Behaviour*, **65**, 11–17.

Bateson, P. P. G. & Klopfer, P. H. 1989. *Perspectives in Ethology, 8*. New York: Plenum.

Berry, M. 1993. *SVDPACKC (Version 1.0) User's Guide*. Cookeville: Tennessee Tech University. Available from: http://www.netlib.org/svdpack/index.html.

Börner, K. 2004. *Info Vis CyberInfrastructure: a Data-Code-Computer Resource for Research and Education in Information Visualization* See: http://iv.slis.indiana.edu/.

Börner, K. & Zhou, Y. 2001. A software repository for education and research in information visualization. Information Visualisation Conference, London, England, July 25–27, pp. 257–262.

Boyack, K. W. 2004. Mapping knowledge domains: characterizing PNAS. *Proceedings of the National Academy of Sciences, U.S.A.*, **101**, 5192–5199.

Boyack, K. W. & Börner, K. 2003. Indicator-assisted evaluation and funding of research: visualizing the influence of grants on the number and citation counts of research. *Journal of the American Society of Information Science and Technology*, **54**, 447–461.

Cooper, B. G. & Goller, F. 2004. Multimodal signals: enhancement and constraint of song motor patterns by visual display. *Science*, **303**, 544–546.

Darwin, C. 1872. *The Expression of the Emotions in Man and the Animals*. London: J. Murray.

Dawkins, M. S. 1989. The future of ethology: how many legs are we standing on? In: *Perspectives in Ethology, 8* (Ed. by P. P. G. Bateson & P. H. Klopfer), pp. 47–54. New York: Plenum.

Dewsbury, D. A. 1989. A brief history of the study of animal behavior in North America. In: *Perspectives in Ethology, 8* (Ed. by P. P. G. Bateson & P. H. Klopfer), pp. 85–122. New York: Plenum.

Dewsbury, D. A. 1992. On the problems studied in ethology, comparative psychology, and animal behavior. *Ethology*, **92**, 89–107.

Dewsbury, D. A. 1994. On the utility of the proximate–ultimate distinction in the study of animal behavior. *Ethology*, **96**, 63–68.

Durant, J. R. 1986. The making of ethology: the association for the study of animal behaviour, 1936–1986. *Animal Behaviour*, **34**, 1601–1616.

von Frisch, K. 1953. *The Dancing Bees*. New York: Harcourt Brace Jovanovich.

von Frisch, K. 1967. *The Dancing Language and Orientation of Bees*. Cambridge, Massachusetts: Harvard University Press.

Gittleman, J. L. 1989. The comparative approach in ethology: aims and expectations. In: *Perspectives in Ethology, 8* (Ed. by P. P. G. Bateson & P. H. Klopfer), pp. 55–83. New York: Plenum.

Griffiths, T. L. & Steyvers, M. 2004. Finding scientific topics. *Proceedings of the National Academy of Sciences, U.S.A.*, **101**, 5228–5235.

Huntingford, F. A. 2003. A history of animal behaviour by a partial, ignorant and prejudiced ethologist. *Animal Behaviour*, **66**, 409–415.

Huxley, J. S. 1942. *Evolution: the Modern Synthesis*. New York: Harper.

Kamada, T. & Kawai, S. 1989. An algorithm for drawing general undirected graphs. *Information Processing Letters*, **31**, 7–15.

Kleinberg, J. 2002. Bursty and hierarchical structure in streams. In: *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1–25. New York: ACM Press.

Klopfer, P. H. & Polemics, J. 2002. The state of ethology. *Animal Behavior Society Newsletter*, **47** (3), 4–5.

Landauer, T. K., Foltz, P. W. & Laham, D. 1998. An introduction to latent semantic analysis. *Discourse Processes*, **25**, 259–284.

Landgraf, R., Frank, E., Aldag, J. M., Neumann, I. D., Sharer, C. A., Ren, X., Terwilliger, E. F., Niwa, M., Wigger, A. & Young, L. J. 2003. Viral vector-mediated gene transfer of the vole V1a vasopressin receptor in the rat septum: improved social discrimination and active social behaviour. *European Journal of Neuroscience*, **18**, 403–411.

Lorenz, K. Z. 1965. *Evolution and Modification of Behavior*. Chicago: University of Chicago Press.

McGregor, P. K. 1992. *Playback and Studies of Animal Communication*. New York: Plenum.

Mane, K. K. & Börner, K. 2004. Mapping topics and topic bursts in PNAS. *Proceedings of the National Academy of Sciences, U.S.A.*, **101**, 5287–5290.

Martins, E. P. 1996. *Phylogenies and the Comparative Method in Animal Behavior*. New York: Oxford University Press.

Martins, E. P. & Clark, A. V. 2002. EthoSource: a global initiative for integrating and sharing behavioural information: http://mypage.iu.edu/~emartins/Ethosource/.

Maynard Smith, J. & Price, G. R. 1973. The logic of animal conflict. *Nature*, **246**, 15–18.

Ord, T. J. & Evans, C. S. 2002. Interactive video playback and opponent assessment in lizards. *Behavioural Processes*, **59**, 55–65.

Phelps, S. M. & Young, L. J. 2003. Extraordinary diversity in vasopressin (V1a) receptor distributions among wild prairie voles (*Microtus ochrogaster*): patterns of variation and covariation. *Journal of Comparative Neurology*, **466**, 564–576.

Salvador, A., Serrano, M. A. & Gonzalez-Bono, E. 2003. Research trends in the journal Hormones and Behavior (1987–2000). *Hormones and Behavior*, **43**, 375–380.

Sherman, P. W. 1988. The levels of analysis. *Animal Behaviour*, **36**, 616–619.

Slater, P. J. B. 2003. Fifty years of bird song research: a case study in animal behaviour. *Animal Behaviour*, **65**, 633–639.

Stamps, J. A. 2003. Behavioural processes affecting development: Tinbergen's fourth question comes of age. *Animal Behaviour*, **66**, 1–13.

Stephens, D. W. & Krebs, J. R. 1986. *Foraging Theory*. Princeton, New Jersey: Princeton University Press.

Thorpe, W. H. 1956. *Learning and Instinct in Animals*. London: Methuen.

Tinbergen, N. 1951. *The Study of Instinct*. Oxford: Clarendon.

Tinbergen, N. 1963. On aims and methods of ethology. *Zeitschrift für Tierpsychologie*, **20**, 410–433.

West, M. J., King, A. P. & White, D. J. 2003. The case for developmental ecology. *Animal Behaviour*, **66**, 617–622.

White, H. D., Lin, X., Buzydlowski, J. W. & Chen, C. 2004. User-controlled mapping of significant literatures. *Proceedings of the National Academy of Sciences, U.S.A.*, **101**, 5297–5302.

Wilson, E. O. 1975. *Sociobiology: the New Synthesis*. Cambridge, Massachusetts: Harvard University Press.

Analysis and Visualization of Animal Behavior dataset

This page describes various tools and practices used for the analysis and mapping the domain of Animal Behavior studies. Links to results and source code are available for download.

Index

I. Authors and Affiliations

- Sidharth Thakur, Computer Science Department, Indiana University Bloomington (IUB)
- Ketan Mane, School of Library and Information Science, IUB
- Katy Borner, School of Library and Information Science, IUB
- Terry Ord, Biology Department, IUB (Currently at University of Cal., Davis)
- Emilia Martins, Biology Department, IUB

II. Introduction

Summary

This work describes application of latest analysis and visualization techniques to the domain of Animal Behavior studies. The motivation emerged from the fact that the field has remained uncharted and only been probed using conventional reviews and smaller subset of published research. We demonstrate that using knowledge-domain visualization techniques can reveal interesting details about the prominent and emerging research areas over the years.

Overview

Traditional methods of mapping and understanding growth of scientific research and literature in the past have usually lacked comprehensive coverage and have been quite painstaking. This changed considerably with the advent of electronic databases and easy access to published research on the world wide web. Particularly, journal citation related information can now provide valuable information for finding associations within a field.

Information gleaned from such electronic records is very helpful in mapping domains and charting various relationships that evolve over a period of time. These associations provide a view of the evolution of the field in the past as well as, emergence and growth of promising areas and opportunities for collaboration. In order to demonstrate the potential of these latest analysis and visualization techniques, we used citation records for the domain focusing on animal behavior studies. The fact that the field has rich publications and has never been mapped in this way provided ample motivation for the study.

Procedure: Our analysis was based on published records belonging to a set of core journals derived from the Biological Abstracts database. These journal records were analyzed using techniques like Latent Semantic Analysis (LSA) and Co-word analysis for representative time slices. Latent semantic analysis was applied to determine semantic similarity among the documents based on their keywords. Geographical information was used to obtain interesting results regarding hotspots of published and ongoing research.

Outcome of study: Areas that emerged corresponded to the leading research fields for all the journals. Further, co-word analysis gave us the changes in topics covered in the years under consideration. Graph visualization software, *Pajek* [1], provided interesting visual insights into relationships within document and keyword spaces. These analytical and visual tools helped domain experts in identifying the focus areas of each journal, their coverage and changes in their dynamic field and have established a foothold for deeper exploration and mapping of the domain.

Initial work was presented as a poster during the Conference of Visualization and Data Analysis, 2004 (view paper here). Subsequently, a paper titled "*Trends in animal behaviour research (1968–2002): ethoinformatics and the mining of library databases*" has been published in Animal Behavior (journal of animal behavior studies; view paper here).

Source: The download directory is located here. This directory contains source code and data pertaining to the analysis.

III. Tools used

We employed several statistical techniques for the work. Specific algorithms and tools were used to analyze the results as graph visualizations. The various techniques that were used are described below along with a brief mention of tools they used:

- **Data extraction** - journal citation records were Data was extracted and cleaned using downloading tools and software parsers described in the following sections
- **Trends in journal publications** - Various Perl parsers were used to obtain basic statistics and trends
- **Semantic analysis of document space** - Latent Semantic Analysis (LSA) algorithm was used to analyze documents for similarities based on words obtained from keywords and title fields of records
- **Co-word semantic analysis** - was carried out by building co-word matrix after extracting the unique keywords. Perl parsers were found to be very efficient and fast in such tasks
- **Bursty behavior of keywords** - An algorithm to determine significant 'bursts' in occurrence of keywords was used to obtain rise in popularity of keywords over the years. This algorithm provides several options to adjust granularity of the results
- **Visualization package** - semantic space of documents and keywords were analyzed using a network graph-drawing application called *Pajek*. This tool was selected for its scalability, useful graph generating algorithms and ease of use
- **Charts and diagrams** - Microsoft Excel was used to analyze and visualize the various trends and graphs

---

IV. Dataset

Description: The primary source of the journal citation records was the Biological Abstract dataset, which consists of about 200 journals grouped into three broad categories. The categories are derived on the basis of classification of Animal Behavior Abstracts at the Cambridge Scientific Abstracts website. They include journals categorized as: Core, Priority and Selective set of journals. The Animal Behavior Abstracts page links to the complete source serial list of the abstracts.

Procedure: Using International Standard Serial Numbers (ISSN) to identify journals, we downloaded all available citation records as ASCII text files from the Biological Abstracts database (BIOSIS, Inc 2002). The standard library web interface prevents data streaming by restricting the number of records that can be downloaded at a time, and can result in a large number of network timeouts and failures. Instead, we used access client software *SilverPlatter WinSPIRS v4.01* (Ovid Technologies, Inc 1999) to download records directly from the server of Indiana University's library database.

Several Perl parsers were used to extract relevant data from downloaded files. Parsers are available on this page via links. Using these parsers, we converted downloaded files to text files (delimited by the pipe symbol or '|') and removed duplicate documents (identified by the string 'title#year#author#keywords'). To verify that our dataset was complete, we noted the number of documents listed by Biological Abstracts during each ISSN search and later crosschecked this number with those files actually downloaded. (Please refer section on Data extraction and cleaning for link to parsers).

Finally, we calculated summary information on the number of unique records for each journal in each year. Publication records for several journals were incomplete, ranging from approximately one year (e.g., Behavioural Processes - 1977) to thirteen years (e.g., Behavioral Ecology and Sociobiology - 1977-1982, 1985-1991; Table 1). Before proceeding with our analyses, we confirmed these records were missing from Biological Abstracts (E. Ten Have, BIOSIS, pers. comm.) and not the consequence of errors accumulated during data acquisition. Our final dataset covered 25 journals from 1968 to 2002 totaling 42,836 records of published material (Table 1; Fig. 1).

Results: In all, there are about half-million records, the bulk of which span the years 1968 through 2002. The downloaded, unparsed data was organized as follows:

- ASCII text files: 340 in all
- Average numb of records: 1800
- Total number of records:
  - about 466,000 (including duplicate records)
  - about 443,000 (unique records only)

The journals were classified into three categories:

1. Core set: 14 journals; 15282 records. Some of the core journals are: *Anim. Behav., Anim. Learn. Behav., Appl. Anim. Behav. Sci., Behaviour, Behav. Ecology, Behav. Ecology and Sociobiology, Behav. Processes, Bird Behav., Birds N. America, Ethology, Journal of Ethology, Journal of Experimental Psychology: Anim. Behav. Processes, Journal of Insect Behav., Learning and Motivation* (subsequently, a set of 25 journals was also used as core-journals, see the statistical analysis section below for the listing)
2. Priority set: 38 journals
3. Selective set: 166 journals

The complete list of journals as categorized using the above is available at this link (text file, 8 Kb).

Format of a typical Record: The downloaded records primarily contain journal citation information that corresponds to the ISI (Institute for Scientific Information) data format. Prominent fields include: Title, Keywords, Author(s), Affiliation (addresses), Abstract, Species taxonomic information, publication Year, etc. among forty fields.

The complete description of most of the fields occurring in the data set is available at this link (html).

A typical record from the original downloaded journal-data is described next. Records are organized by serial numbers and the record header includes information about the source in the Biological Abstracts database. For example, the following is a record header -

Record 1 of 29 - Biological Abstracts 2001/07-2001/09

Record header is followed by the citation details namely, Title, Author, Address, Source or Journal name, Publication year, and so on. This data is in the following format -

Table 1:

Field-name Data

TI: Additional material of the enigmatic golden mole Cryptochloris zyli, with notes on ...
AU: Helgen-K-M {a}; Wilson-D-E
AD: {a} Mammal Department, Museum of Comparative Zoology, Harvard University, Cambridge, MA, 02138: helgen@fas.harvard.edu, USA
SO: African-Zoology. [print] April, 2001; 36 (1): 10-112.
PY: 2001
…

There were several issues with the format and availability of the main fields within the records. The description of the main fields and issues pertaining to their extraction are available at this link (html file).

V. Data Extraction and Cleaning

Description: The analysis focused on the vocabulary chosen for titles and keywords of animal behavior publications. We used title-words to explore major changes in topics and study organisms of interest across the three decades included in our dataset. Keywords contained more detailed information and had the advantage of retaining compound terms, such as 'sexual selection' and 'parental care'. We used keywords for more detailed analysis of journal and geographic trends.

Several issues existed with the raw data which had to be addressed before proceeding with the analysis. These are mentioned here -

1. Missing Data - The quality of downloaded data was dependent on the active sessions while using the Winspire software. Some data might have been lost if the

connection with the BA database broke prematurely. This led to some missing data or gaps in the dataset which were filled at a later stage.
2. Lack of cited references - the data set did not include the cited references. The analysis did not focus on cited references.
3. Duplicate records - duplicate records existed for several entries and these had to be eliminated before analyses could be carried out.
4. Missing entries - most records did not contain all the forty fields mentioned in the links above. Further, there were certain fields like the URL that occurred only in a very small percentage of the records.
5. Split files - journals that had very large number of records were split into smaller files. These files were combined into a single file when extracting data for individual journals.

Procedure: In our analysis of title-words, we began by using parsers to remove uninformative words such as 'of' and 'the', identified from stop-word lists (available from Börner & Zhou 2001; Börner 2004). Keyword information is presented in Biological Abstracts as a string of semi-colon separated compound words (e.g., "anthophilous insects; breeding systems; climate severity; disturbance; evolution; habitat") and are found in two separate fields: DE or 'descriptors', and MI or 'misc. indicators'. We combined keywords from DE and MI fields before conducting further analysis. For both title and keyword analyses, we also removed "behavio/ur" and "behavio/ural", which occurred often enough that they might obscure more subtle document associations.

Data was extracted from the ISI-formatted files and extracted in the following steps:

1. The pipe-character ('|') was used as the record delimiter.
2. The parser for `data-extraction (perl file)` was employed to obtain the records from all the 340 files into a single '|'-delimited files
3. The parser for `duplicate-removal (perl file)` was used to extract the unique records. Records were identified as duplicate if they contained the same string from the concatenation of following fields (ISI acronyms)
    1. Title (TI)
    2. Year (PY)
    3. Author (AU)
    4. Descriptors (DE) - keywords
    5. Miscellaneous Descriptors (MI) - keywords
4. All common and uninformative words were eliminated. The list of such stop-words is available at this link (text file)
5. All the text characters were converted to upper-case to maintain uniformity.

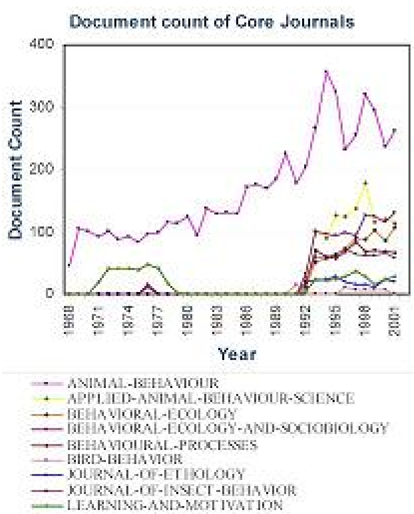
VI. Statistical Analysis

Description: We began our analysis by ascertaining the trends in volume of publications over the years spanned by the available records. We hoped to gain valuable information regarding spans of year that could be of particular interest.

Procedure: Unique records were isolated and counted for individual journal and all the journals respectively to generate the distributions discussed below.

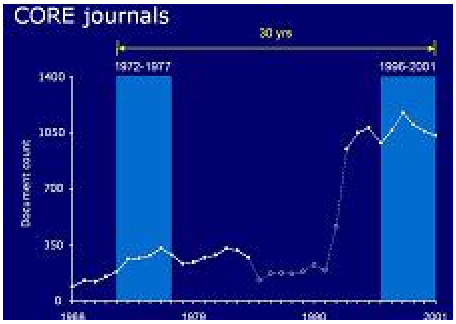
Results:

Set of 13 core journals - The initial data set consisted of records from 1968-2002. The number of records for various years for the following journals selected from the core set of journals is shown below:
Anim. Behav., Appl. Anim. Behav. SCI, Behav. Ecology, Behav. Ecology and Sociobiology, Behav. Processes, Journal of Ethology, Journal of, Journal of Insect Behav., Learning and Motivation



(click to enlarge) Figure 1: Number of journal publications for journals selected from the set of core journals

The following figure shows the trends in core journal group. Two groups of years were focused upon namely, 1972-77 and 1996-2001. The blue bars correspond to these two groups: :



(click to enlarge) Figure 2: Trends in publication in select core journals within Animal Behavior journals

The above figures show the growth of the Animal Behavior domain over the years.

The initial statistics focused on gathering information about the extent of the dataset and the possible fields that could be of interest. The data gathered is presented below:

1. Publication Years: 43, spanning 1915-2002 (only partial data for the year 2003 was available)
2. Titles: about 443,000 records that had Titles
3. Keywords: 103,000 DE terms and 600,000 MI terms
4. Authors: about 537,000 entries (duplicate or alternate author names were not removed)
5. Journal Source (SO): about 240
6. Institution addresses (AD): about 296,000 (geographical distribution was analyzed in detail later)

Parsers like `get-unique_years (perl file)` and `get-frequency-yearwise (perl file)` were used to determine the frequency of records pertaining to the above fields.

As a first analysis, journal articles published in three years namely, 1994, 1997 and 2000 were selected. This was done to uncover the pattern of growth over the recent decade and have a manageable dataset to work with initially. The motivation was to discover whether the techniques could faithfully display the trends of growth in the field as compared to the known trends in the domain. The following table presents the year-wise distribution of documents, keywords, etc.:

Table 2:

Year	**1994**	**1997**	**2000**
Number of documents	648	778	740
Number of unique keywords	1244	2324	2269
Average number of unique keywords	1.92	2.99	3.06

Several documents having zero or one keyword were excluded from the above set of documents. It can be observed that the average number of keywords increases over time while the number of paper and the number of unique keywords roughly stay the same.

The gaps mentioned in the above figures were filled by downloading latest data and appending to the dataset. With a more complete dataset, analysis was carried out again. For the purpose of our analysis, the journals were divided into four groups. The groups comprised of journals selected from particular year-spans. The groups and the number of unique records available for these groups are as follows:

Table 3:

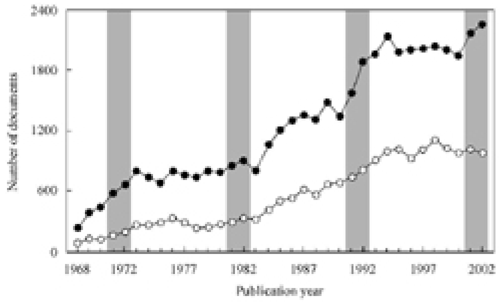
Year span (Groups)	**2000-2002**	**1990-1992**	**1980-1982**	**1970-1972**
Number of documents	5668	4223	2681	1345

The parser get-records-batchwise (perl file) was used to extract the data into the above mentioned groups.

Results:
Set of 25 core journals - The core journals were further organized into a set having 13 of the original core journals (see listing in Datasets above) and a more comprehensive set of 25 journals. The set of 25 journals that were selected into the core set were as follows:

Journal of comparative psychology, Learning & Memory, Behavioral Ecology, Journal of Insect Behavior, Behavior Research Methods, Behavioral Neuroscience, Ethology Ecology & Evolution, Behavioural Processes, Behavioral Ecology and Sociobiology, Journal of Ethology, Ethology, Applied Animal Behaviour Science, Behavioural Brain Research, Bird Behavior, Behavioral and Brain Sciences, Journal of Experimental Psychology: Animal Behavior Processes, Aggressive Behavior, Animal Learning & Behavior, Physiology & Behavior, Learning and Motivation, Journal of the Experimental Analysis of Behavior, Hormones and Behavior, Behaviour, Animal Behaviour, Behavior Genetics.

Our final dataset covered 25 journals from 1968 to 2002 totaling 42,836 records of published material, see figure .



(click to enlarge) Figure 3: Trends in publications in the larger set of core Animal Behavior journals (black dots).
White dots indicate journal count from the set of 13 journals after database was updated for incomplete records.

The number of documents for each year for the journals in the two sets (set of 13 and 25 journals) were determined. The results are linked below:

1. Year-wise document-frequency for set of 13 journals in the core set at this link
2. Year-wise document-frequency for set of 25 journals in the core set at this link

---

VII. Analysis of Journal Coverage

Description: Journal coverage was determined by observing the terms occurring with high frequencies.

Preliminary observations -

Procedure: The words in bold in the above table represent the top keyword for the corresponding journal. The parser `get-top-words-journal-wise (perl file)` was used to determine the distribution of terms for individual journals.

Results: The first part of the domain analysis focused on identifying the major keywords within the core journals. The top ten keywords based on frequency of occurrence were obtained for the following core journals:

Table 4:

Journal	Keywords
Journal Of Animal Behavior	Behavior, Aggression, Animal-Behavior, **Evolution**, Male, Female, **Sexual Selection**, Body-Size, **Foraging**, Reproductive-Success, Mate-Choice
Journal Of Bird Behavior	**Aggression**, Adult, Feeding, Male, **Reproductive Success**, Species Interaction,

	Vocalization, Brood Parasitism, Brooding, Brood Mates
Journal Of Behavioral Ecology	Behavior, **Sexual Selection**, Body Size, Predation Risk, Reproductive Success, Mate-Choice, Fitness, Parental Care, Male, Reproduction
Journal Of Learning And Motivation	Behavior, **Learning**, Neural Coordination, Conditioning, Rat, Conditioned Stimulus, **Motivation**, Pavlovian-Conditioning
Journal Of Applied Animal Behavior Science	Behavior, Animal Welfare, **Animal-Behavior**, Animal Husbandry, Stress, Aggression, Housing, Meeting-Abstract, Social-Behavior, Grazing, Feeding
Journal Of Insect Behavior	Oviposition, Female, **Foraging-Behavior**, Body-Size, **Sexual Selection**, **Foraging**, Reproduction, Male, **Parasitoid**
Journal Of Behavioral Process	Behavior, **Learning**, Abstract, Animal-Behavior, Nervous System, Reinforcement, Social-Behavior, Memory, Female, Foraging-Behavior, Aggression
Journal Of Behavioral Ecology And Sociobiology	**Sexual Selection**, Reproductive-Success, Body-Size, Evolution, Female, Aggression, Male, Competition, Reproduction, Sperm Competition
Journal Of Ethology	Female, **Aggression**, Body-Size, Male, **Dominance**, Copulation, Spawning, Foraging, **Mating-Behavior**, Social Behavior

Further, top words for the three years (1994, 1997, 2000) were also determined likewise.

Table 5

Year	Keywords
1994	Behavior, Animal-Behavior, **Animal-Communication**, **Evolution**, Mathematical-Model, Aggression, **Foraging**, Predation, Seasonality, Learning
1997	Behavior, Male, Female, **Reproduction**, Ecology, Adult, Sexual Selection, Animal Husbandry, **Evolution**
2000	**Sexual Selection**, Body-Size, Aggression, Reproductive-Success, Animal-Welfare, Predation-Risk, Territoriality, Mating-System,

Mate-Choice, Competition

The words in bold represent the top keywords for the corresponding year group. The parser `get-top-words-year-wise (perl file)` was used to determine the above distribution.

Journal coverage of wider scope of journals -

Procedure: Word frequencies were determined by isolating unique terms from the keyword fields. For each of the 25 journals, the top ten most frequently occurring words were extracted. A total of 143 different terms were found and only a moderate degree of overlap across the journals was observed.

Results: When the most common keywords for all journals are pooled and ranked, terms referring to aggression, learning/memory, foraging and sexual selection top the list, see figure .

Journal keywords also show the existence of a continuum between serials that emphasize evolutionary ethology and those that publish comparative psychology. On one extreme, Behavioral Ecology and Behavioral Ecology & Sociobiology place an unusually strong emphasis on "sexual selection" and other terms related to reproduction and mating, see figure . Animal Behaviour, Behaviour and Ethology also publish articles on sexual selection, but add social behaviour, predation, foraging, communication (usually "vocalization"), and evolution. At the other extreme, nearly all popular keywords reported by Behavioural Processes and Animal Learning & Behavior address animal learning and memory, see figure .

Interestingly, Applied Animal Behaviour Science lies somewhere in between, including "aggression" and "vocalization" as prominent keywords, but also "stress" and "motivation". Most of the journals that appear on our supplemented list of 25, but not in the core list from Animal Behaviour Abstracts, also appear somewhere in the middle. For example, physiology journals like Hormones & Behavior frequently publish studies about "sexual behaviour" and "aggression", in addition to "stress", "learning" and "photoperiod".

The Journal of Comparative Psychology shows an unusually high peak with "animal communication" appearing along side research relating to "habituation", "development" and "imitative learning". These journals therefore form an important bridge between the academic descendants of early ethologists and comparative psychologists. They also tend to be more specialized than journals at the extremes, leading to a more highly skewed distribution of keywords and/or a larger proportion of unique terms, see figure . For example, the focus of Applied Animal Behaviour Science is on "animal welfare", whereas Hormones and Behaviour publishes more papers described by "neuro/endocrinology".

VIII. Analysis of Semantic Document Space

Description: The semantic similarity between the documents representing the records of the journals were determined. Latent semantic analysis (LSA) [2], also called latent semantic indexing, was applied to determine semantic similarity among the documents based on their keywords. LSA extends the vector space model by modeling term-document relationships using a reduced approximation for the column and row space computed by the singular value decomposition of the term by document matrix. The strength of LSA lies in resolving the fundamental issues concerning the conventional lexical matching schemes namely, synonymy (similar meaning words) and polysemy (words with multiple meaning) [3].

Document-Term analysis refers to the application of LSA on the term-by-document (TD) matrix formed from the documents and their terms. The TD matrix has the following format:

1. Columns are document number or id
2. Rows are individual terms
3. Cells of the matrix contain the number of times the particular term (say k-th term) occurs in a document (say m-th document)

Table 6: Term-Document matrix

m	Doc 1	Doc 2	...	Doc M
k				
Term 1	2	0		
Term 2	1	2		
:				
Term K				

Singular Value Decomposition (SVD) is a vectors-based model that is used to determine the important latent dimensions. SVD analysis is performed using the LSA SVDPACKC provided by M. Berry [2]. The SVDPACKC takes as input a similarity matrix in the format specified by the Harwell-Boeing format (hbf). Intermediate files provide important latent dimensions. Output of the pack is a similarity matrix that specifies similarity between the row and column entries. These values are normalized values (matrix values divided by the highest value).

Procedure: Data parsing, generation of unique terms and term vs. document frequency matrices, and similarity matrix computations were carried out using code available in the Information Visualization Repository at Indiana University. The LSA SVDPACKC provided by M. Berry [4] was applied to determine the most important latent dimensions. The most significant dimensions obtained for the three years are: 1994 (114 dimensions), 1997 (112 dimensions) and 2000 (114 dimensions).

Visualizations of the semantic relationships among similar documents were generated using the Pajek graph visualization software [1]. The Kamada Kawai algorithm [5] implemented in Pajek was used to layout the documents in a 2-dimensional space.

The parser `get-records-batchwise (perl file)` was used to extract the keywords from the DE and MI fields.

The semantic analysis was carried out in the following steps:

1. Extract keywords into a semicolon (';') delimited file
2. A Java parser was used to convert the data into a sparse format known as Harwell-Boeing format
3. LSA was applied using the Java package available in SVDPACK to obtain a document-keyword similarity matrix
4. The similarity matrix was converted into Pajek format file using the parse `get-keydoc-to-pajek-format (perl file)`
5. The Pajek format input file was viewed using the Pajek browser
6. Within the browser, the Kamada-Kawai algorithm was used to obtain the clusters of various documents represented by nodes. Nodes were connected if they were similar based on the keywords as determined by LSA
7. Initially, owing to very high similarity, the nodes were highly interconnected and resulted in a highly dense network of associations. In order to view only the most prominent clusters of nodes (documents), a weight or threshold value was used to eliminate edges below this threshold-value. This resulted in clusters of nodes

Results: In the 1994 dataset, three clusters were identified, see figure. The first cluster (blue background) deals with documents dealing with parental behavior. The second cluster in red covers animal behavior and learning research. The gray cluster contains documents on feeding behavior.
The 1997 dataset contains three main clusters, containing documents on aggressive social behavior, sexual selection and mating behavior and sexual behavior see figure.
In 2000 only two clusters were identified: mating and foraging behavior as well as nesting behavior, see figure.

The document space was also analyzed using the words derived from the title (TI) field of the records. Titles-words were extracted using the parser `get-title-keywords (perl file)`. Titles include several commonly used words like the articles, etc., which must be excluded from the analysis. These words were removed by referring to the list of stop-words, which is available at this link (text file). The title analysis was carried out for two spans of years namely 1972-77 and 1996-2001.

IX. Analysis of Co-word Occurrence Keyword Space

Description: Co-word or Term-Term or co-Term analysis focuses on similarity between keyword pairs based on how frequently two or more keywords occur together. Term-Term (TT) similarity matrix is obtained by considering all the unique keywords and the words or terms they appear with in the documents.

The TT matrix has the following format:

1. Column headers are unique Terms or keywords
2. Row headers are unique Terms or keywords
3. Cell values indicate the number of times the k-th term occurs with the m-th term

Table 7: Term-Document matrix

m	Term 1	Term 2	...	Term M
k				
Term 1	2	0		
Term 2	1	2		
:				
Term K				

In order to analyze the change in the topics covered by the journals within the various groups, a keyword-keyword matrix was built using words that occurred together more the once. The similarity matrix was then converted into Pajek format using the parser `get-keydoc-to-pajek-format (perl file)` and visualized within the Pajek browser. The Fruchterman-Reingold 2D-algorithm [6] was employed to visualize the network of keywords. Edges represented associations between keywords that occurred together. A threshold was used to obtain the prominent clusters.

Results: The keyword, BEHAVIOR emerged as the central, highly interconnected node, bridging different characteristics of animals like Aggression, Mating, Welfare in all the three time slices.

For the year 1994, the most dominant areas of study were identified by: Animal Communication, Evolution and Foraging, see figure .

Similarly, for the year 1997, the most dominant areas of study were identified by: Reproduction and Evolution, see figure .

In 2000, the study of Sexual Selection dominated the research. Some of the co-occurring keywords also described the study of Natural Selection, Mating Success, and Courtship, see figure .

X. Burst Analysis

Description: Burst analysis [7] is a technique that aims to analyze documents to find features that have high intensity over finite/limited duration of time periods . Rather than using plain frequencies of the occurrences of words, the algorithm employs a probabilistic automaton whose states correspond to the frequencies of individual words. State transitions correspond to points in time around which the frequency of the word changes significantly. Details of using this algorithm are available at this link (text file) .

Procedure: The terms occurring in the title were used for burst analysis. The terms were isolated using this parser `p_extract_burst_title_kywds (perl file)`. As applied here, the burst detection algorithm focuses on the temporal intervals between repeated appearances of the same term. When a term is popular, it will be used frequently and the time intervals between repeated appearances will be short. The two-state form of the burst detection algorithm finds the model that best describes the data as a collection of temporal strings of high (i.e., bursts) and low episodes of popularity for each of the terms studied. ‘Weights’ are also calculated to allow for direct comparison among bursts for the same and different words in terms of their relative prominence.

Words such as “the” and “an” (and other such words given in this link (text file)) were excluded to obtain a total of 24,850 unique title-words. We further focused our attention on the 739 title-words that appear at least 100 times in the dataset. To these, we applied the burst detection algorithm, looking across the full 35 years of publications, to identify rapid increases and decreases in popularity (bursts) for each term through time.

The input for the burst detection algorithm was generated `p_extract_burst_input (perl file)`. Other parsers were used to generate keyword and year-wise frequencies. These parsers are available in this folder .

Results: the burst detection algorithm identified 506 bursts of popular title-words across 35 years. Bursts were regularly spaced, lasting a median of 4 years (mean±SE =

5.6±0.19). There were 470 title-words such as “effect”, “role” or “difference” that are difficult to interpret further. If we focus on the remaining 269 terms, there appears to be three vocabulary periods: pre-1985, 1985-1995, and post-1995, see figure .

Of the 269 potentially meaningful terms, 200 reflect major topics of interest in animal behavior research (e.g., “operant”, “evolution”, “predation”). The words occurring within each of the three time periods cross disciplinary boundaries, indicating the continuing diversity of animal behavior research throughout the history of our field. The early periods, for example, show bursts from “shock”, “reinforcement”, “natal” and “testosterone”. “Guarding”, “genetic”, “anxiety”, and “opioid” all burst during the transition time period (1985-1995). “Receptor”, “anxiety”, “paternity” and “mate” all burst in the most recent time interval.

The 69 remaining terms refer to a type of animal (e.g., “rats” appears 5,550 times; see figure), and also reflect some meaningful shifts over the years. Before 1985, virtually all animal terms undergoing bursts of popularity are model organisms, including cats, monkeys, squirrels (which could also be ‘squirrel monkeys’), and chickens (Fig. 3b). In the 1985-1995 transition period, there are several bursts referring to insects (especially hymenoptera (e.g., bees, wasps, and ants) and orthoptera (e.g., crickets, grasshoppers and katydids)) that were not abundant earlier. In the mid-1990s, there is a sudden surge of interest in a more diverse group of domesticated animals and animals of economic importance (e.g., dogs, cows, deer).

XI. Geographical and Institutional Patterns

Description: This section of work focused on most current publishing trends across the world. By focusing on contents published in most popular journals of the animal behavior domain, the goal was to identify if there exists any common pattern in the research at a global level. Further the study was extended to identify most active institutions of North America in animal behavior domain.

Procedure: Latest subset of data namely for years 2001-2002 was used for data analysis. Further all publication records were split into two datasets: 13 core-journals and 25 all-journals. Both datasets, one comprising of core-journals (670 records) while the other consisting of all 25 journals (1806 records) were analyzed separately to identify global contribution towards journals. Using five digit zip-code pattern available for North America as markers, the paper coverage in both datasets was initially split into two categories: 1) US and 2) Non-US publications. Excluding records with missing country name abbreviations, all other country names from Non-US dataset were split into seven identifiable regions on a global topology. So for both datasets, eight geographical locations (including North America) identified are shown here,

- North America

- South America
- Western Europe
- Eastern Europe
- Middle East
- Asia
- Africa
- Australia/New Zealand

The 13 core-journals and 25 all journal datasets was individually split into these eight geographical locations. Keywords from these individual regions were used to identify research trends across these eight identified regions. Further zip-code was used as tags to identify US institutions contributing to core-journals and all 25 journals within the animal behavior domain. Frequency count of zip-codes was calculated to determine top 25 US institutions from both the datasets. A similar analysis could not be replicated at a global level, as zip (or post) codes in many countries are less likely to be specific to particular institutions. In addition, variation in abbreviated names of institutional names and its location within the address string proved to be difficult to be isolated for the study.

Parsers were used to isolate records for US-based institutions and obtain information pertaining to specific zip-codes and/or institutions. The parsers are available in this folder. The parser `p_usa_inst.pl(perl file)` was used to segregate records based on Us or Non-Us origin. Zip code frequencies were determined using the parser `p_zipCodeFreq.pl (perl file)`.

Results:
Region-wise distribution: In 2001–2002, animal behavior publications were produced by researchers affiliated with institutions in every region of the world, see figure, with North America and Western Europe being the primary producers of animal behavior research. Keywords used by North America, Western Europe and Australia/New Zealand were remarkably similar, reflecting global agreement on popular topic areas in animal behavior, see figure.

The relatively few contributions from remaining regions (South America, Eastern Europe, Asia and Africa) shared an emphasis on animal learning and domesticated animals, indicating strong interest in applied animal behavior research. Relative representation by South America and Asia was larger when considering all 25 journals, whereas Africa, Middle East and Eastern Europe were better represented when only core ABA serials were considered.

Zipcode-wise distribution (US only): In the United States alone, we counted 1806 publications from first authors located at more than 487 different zip codes during the 2001–2002 period. Over 100 zip codes tallied five or more animal behavior publications, although it is possible that some zip codes do not identify unique institutions (e.g. some could be personal residences). Also, the major representation of some institutions may result from an unusually large number of contributions by single investigators in this 2-

year period. The number of documents associated with each zip code also varied between the two lists (13 core journals and complete list of 25). Nevertheless, several institutions known to have larger graduate programs or a greater number of animal behavior researchers account for a disproportionate share of publications by geographical location within the U.S., statistics available at this link.

References

1. Batagelj, V. & A. Mrvar *Pajek: Program Package for Large Network Analysis.* University of Ljubljana. Slovenia

2. Landauer, T.K., Foltz, P.W. And Laham, D. 1998. *Introduction To Latent Semantic Analysis*. Discourse Processes, 25, 259-284
3. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K. And Harshman, R. 1990. *Indexing By Latent Semantic Analysis*. Journal Of The American Society For Information Science, 41 (6), 391-407
4. Berry, M. 1993. Svdpackc (Version 1.0) User's Guide, University Of Tennessee Tech. Available Online At Http://Www.Netlib.Org/Svdpack/Index.Html
5. Kamada, T. And Kawai, S. 1989. An Algorithm For Drawing General Undirected Graphs. Information Processing Letters, 31 (1), 7-15
6. Thomas M. J. Fruchterman And Edward M. Reingold. 1991. Graph Drawing By Force Directed Placement. Software: Practice And Experience, 21, (11)
7. J. Kleinberg. Bursty and Hierarchical Structure in Streams. Proc. 8th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining, 2002