South African Computer Journal
Number 14
July 1995

Suid-Afrikaanse Rekenaartydskrif
Nommer 14
Julie 1995

Computer Science and Information Systems

Rekenaarwetenskap en Inligtingstelsels

The South African Computer Journal
An official publication of the Computer Society of South Africa and the South African Institute of Computer Scientists

Die Suid-Afrikaanse Rekenaartydskrif
'n Amptelike publikasie van die Rekenaarverenging van Suid-Afrika en die Suid-Afrikaanse Instituut vir Rekenaarwetenskaplikes

Editor
Professor Derrick G Kourie
Department of Computer Science
University of Pretoria
Hatfield 0083
Email: dkourie@dos-lan.cs.up.ac.za

Subeditor: Information Systems
Prof Lucas Introna
Department of Informatics
University of Pretoria
Hatfield 0083
Email: lintrona@econ.up.ac.za

Production Editor
Dr Riël Smit
Mosaic Software (Pty) Ltd
P.O.Box 23906
Claremont 7735
Email: gds@mosaic.co.za

World-Wide Web: http://www.mosaic.co.za/sacj/

Editorial Board

Subscriptions

	Annual	Single copy
Southern Africa:	R50,00	R25,00
Elsewhere:	$30,00	$15,00

An additional $15 per year is charged for airmail outside Southern Africa

to be sent to:
Computer Society of South Africa
Box 1714 Halfway House 1685

Phone: +27 (11) 315-1319 Fax: +27 (11) 315-2276

Guest Contribution

About the author

Dr. Matthew Jones from the University of Cambridge (Management Studies Group) is in fact trained in the Agricultural and Environmental Sciences with a PhD from the University of Reading (1984). He currently holds the position of University Lecturer in Information Management at Cambridge.

He was the secretary and treasurer of the British Institute of Energy Economics Energy Modelling Study Group 1984-1992, member of the Cambridge Health Authority Information Strategy Group 1989-1991. He is also currently member of the British Human Computer Interaction Group, member of the DTI Human Interface Club Special Interest Group on the Organisational Aspects of Information Technology and treasurer and board member of HICOM Human Computer Interaction electronic conferencing system.

He is a highly regarded scholar and has published a substantial number of academic papers in scholarly journals. Some of his more recent publications on business process reengineering (the topic of his contribution) are: (a) Jones, M.R. Don't emancipate, exaggerate: rhetoric, reality and reengineering. In: Baskerville, R.L.; Smithson, S.; Ngwenyama, O. & DeGross, J.I. (Eds) Transforming Organizations with Information Technology. North Holland, Amsterdam, pp357-378, (1994). (b) Jones, M.R. Empowerment and Enslavement: Business Process Reengineering and the Transformation of Work. Paper presented at the "New Visions of the Post-Industrial Society" conference. University of Brighton, (9-10 July 1994) (c) Jones, M.R. The Contradictions of Reengineering. Paper presented at the "Management Challenges in IS" Conference, Cranfield University, 12-13 July 1994.

Business Process Reengineering: Management's New Paradigm or Emperor's New Clothes?

Matthew R Jones

University of Cambridge, Management Studies Group, Department of Engineering, Mill Lane, Cambridge CB2 1RX, UK

e-mail: mrj1@uk.ac.cam.eng

Abstract

Business Process Reengineering (BPR) is currently attracting much attention as an approach to organisational change, but has received relatively little critical scrutiny. This paper seeks to examine the origins, antecedents and content of BPR as they are described in the main writings on the topic, in order to address three questions: Is there one form of BPR or many? Is it new? What factors may account for its popularity? Some of the criticisms that have been made of BPR are also discussed.

Keywords: *Business Process Reengineering*

Computing Review Categories: *J.1*

1 Introduction

Business Process Reengineering (BPR) has been described as the "hottest management concept since the quality movement" [8] and as "one of the key management concepts of the 1990s" [23]. Klein [41], for example reports that 88% of a survey of senior executives claimed that they were engaged in reengineering projects, while Cane [10] reports that up to 70% of large UK companies have undertaken, or are about to undertake BPR initiatives. It has certainly also received its fair share of attention in the management literature in recent years. An on-line search using the terms "Business Process Reengineering", "Reengineering", "BPR" and "Business Process Redesign" in early 1995, for example, identified 3,592 journal, magazine and newspaper articles on the topic since 1990.

Yet, despite the huge amount that has been written about it, the concept of BPR remains elusive. Is it simply the latest buzzword, "that people bandy around ... even if they are just putting in a new computer system" as some popular reports suggest [65], or is it the most important contribution to thinking about organisations since Adam Smith's *Wealth of Nations* as Hammer & Champy [29] claim? Is it primarily an extension of traditional time and motion studies as Parker [54] states, or of Total Quality Management as Shillingford [57] describes, or is it an entirely "new paradigm" [26]? And what is the relationship between Information Technology (IT) and BPR? IS BPR simply a way of "rebadging" existing IT products to increase sales [65], or is it one of the resources to be reengineered as Morris & Brandon [51] imply, or should IT be seen as enabling new processes as Coulson-Thomas [11] suggests, or is it the primary enabler as Davenport [13] proposes?

The confusion surrounding these questions suggests the need for a critical review of the concept of BPR; to examine its origins, antecedents and content. From this it is hoped to develop a clearer picture of the nature of BPR. Is there only one type of BPR or are there several alternatives? What is the substantive content of BPR and how much of this is new? What factors may account for the interest shown in it? This paper therefore seeks to complement the small number of critical analyses of BPR, such as Grint [25], Willmott [67] and Grey & Mitev [24] that have recently begun to emerge. It differs from them however in focusing primarily on the "theory" of BPR, such as it is, rather than seeking to locate the concept within broader social and organisational changes.

For practical reasons it should be clear that this paper cannot aim to provide a full literature review. Moreover a large proportion of the articles consist of reports (frequently quite brief) on the BPR phenomenon for different audiences, or "case studies" of BPR success (often written by the consultants involved), rather than discussions of the nature of the concept. This paper therefore concentrates on those publications widely cited as influencing the development of BPR, or offering a more extended discussion or substantial report on it.

2 Origins of BPR

A number of writers have sought to suggest that the origins of BPR go back many years. Holtham [34], for example suggests that the concept of 'breakthrough' management coined by Juran [39] is "virtually the same" as reengineering, while Bartram [3] traces the "theoretical background of reengineering" to a 1961 Harvard Business School paper on work simplification. Klein [42] even identifies reengineering projects in the Nineteenth Century and cites Henry Ford as an early exponent.

Whatever the merits of these arguments, the evidence of the on-line search of bibliographic databases indicates that the vast majority of the literature on BPR has appeared in the last five years, with 96% being published since 1992. This may be attributed to the influence of two articles published in the latter half of 1990 in the *Harvard Business Review* and *Sloan Management Review*. Both discussed the need for companies to reorganise along process lines, exploiting the opportunities provided by developments in Information Technology.

The *Harvard Business Review* article, "Reengineering Work: Don't Automate, Obliterate" by Michael Hammer [28], a former computer science professor at MIT, is the more widely cited and is credited by many with starting the BPR bandwagon rolling. It introduced the term "reengineering" and set the tone for much later writing on the topic. The rather more academic article in the Sloan Management Review by Thomas Davenport & James Short entitled "The New Industrial Engineering: Information Technology and Business Process Redesign" [14], was associated with the MIT "Management in the 90s" programme [56]. This introduced the acronym BPR and also, as the title indicates, explicitly linked it with IT. The ideas in this paper were later extended by Short & Venkatraman [58] to address processes operating across the boundaries of the traditional business firm, an activity they called Business Network Redesign.

Although Hammer [28] does not actually use the term "Business Process Reengineering", and indeed has recently criticised its widespread use in association with his ideas [30], when he described reengineering as "the radical redesign of business processes" he was clearly talking about a similar activity to that of Davenport & Short. The two papers also referred to some of the same examples as illustrations of their concepts. Thus, even if reengineering and business process redesign are not completely synonymous they would appear to be sufficiently closely related (in practice, if not necessarily in principle) that the use of BPR as an acronym to describe them both would seem justified. On similar grounds, therefore, BPR will, for the purposes of this paper, also be used to describe the various process-oriented approaches to significant business re-organisation that have emerged since 1990, even though their authors may originally have employed alternative names such as Business Process Improvement, Business Reconfiguration, Business Reengineering, Core Process Redesign, Corporate Reengineering, or Process Innovation.

3 Antecedents of BPR

In his original *Harvard Business Review* article [28], Hammer did not seek to associate BPR with any theory of organisations, indeed he presented it specifically as overthrowing arrangements of work organisation that he traced back to the Industrial Revolution. Instead, his argument was based pragmatically on the experience of process redesign in two companies. Although the argument and examples were later elaborated in Hammer & Champy [29], this also does not attempt to develop a theoretical basis for BPR. Rather it was suggested that it is a "conceptually new business model" with no precursors.

Davenport & Short [14], in contrast, described BPR as "the new industrial engineering (IE)". Earlier IE, they explain, was based on F.W. Taylor's extremely successful "mechanizing vision". This was effective in the past, they argue, because the business environment was "largely stable". In today's turbulent business environment, however, a new approach is needed. It is not clear, though, whether Davenport & Short see BPR as rejecting, transcending, or pragmatically adapting Taylorism, although it would appear that they certainly wish it to be seen as lying within the tradition of the "scientific school of management". This view would seem to be supported by Parker [54] who describes BPR as involving time and motion techniques, by Morris & Brandon [51] who associate it with "IE, time and motion, Operational Research and Systems analysis" and by Glover (quoted in [65]) who suggests that it is "upmarket O & M". Wheatley is quoted in Brown [6] as describing BPR as the "supernova" of this "mechanical view of organisations".

Although Hammer and Davenport & Short are generally acknowledged as the progenitors of BPR, it does not mean that their view of the topic can necessarily be seen as definitive. BPR is a commercial product as much as a theory of work organisation. For the customer of BPR, therefore, it is the approach they buy which is likely to define it for them, even if this does not conform to Hammer or Davenport & Short's model. Thus, for example, despite Hammer & Champy 's emphatic statement that Reengineering is not "the same as quality improvement, total quality management (TQM), or any other manifestation of the contemporary quality movement" [29], Shillingford [57] describes BPR as "often evolving out of a TQM initiative" and suggests that its rationale is that "concepts like Japan's 'best manufacturing practice', which uses Just-in-time techniques, can be applied to the office as well as the factory". This view is not a simple misunderstanding, but is based on the approach to BPR developed at Rank Xerox, who were selling it, with accompanying software tools, as a consultancy service (at a starting price of $100,000) [22]. The association of BPR with TQM is also made by Johannson *et al* [36] and *inter alia* by Davenport [13], while Earl [19] and Mumford [52] link it with lean production [68].

Other writers have identified BPR with a number of different organisational and technical developments. For example Coulson-Thomas [11] links it to the emergence of the "network organisation", Stewart [59] to the "horizontal organisation" [53] and Kaplan & Murdock [40] to Dichter's vision of the organisation of the 90s [17]. More bluntly, Byrne [8] suggests that BPR can sometimes be seen as "little more than a euphemism for laying people off", and it is certainly recognised in the popular BPR literature that part of the interest in it comes from companies seeking to reduce staff numbers in the face of economic recession [65]. In some reports BPR is also linked with the implementation of specific types of technology, usually image processing and workflow systems for office automation [57].

Given that the theoretical antecedents of BPR may be expected to be reflected in the approach adopted, i.e. those seeing it as an extension of TQM may be expected to draw on techniques and social structures historically associated with Quality programmes, then this diversity of views on the source of BPR thinking could suggest that there are several distinct versions of it. Another possibility is suggested by Davenport [13] who argues that Process Innovation, as he calls it, does not have a single source, but has developed from the combination of five different approaches to business improvement. He identifies these precursors as: the quality movement, industrial engineering (and its offshoot the soft systems movement *sic*), the socio-technical school, the technology transfer literature and the IT for competitive advantage literature. Even if it is accepted that these approaches can practically be combined, however, it is possible to question whether such a mixed pedigree would seem any more likely to yield a consistent approach to BPR than the various different precursors on their own.

As has already been noted, however, BPR is not just a theoretical construct, but a guide to practical organisational intervention. Whether BPR's theoretical antecedents are consistent, or even what they are thought to be, may not matter in practice, therefore, if what is actually advocated is the same for all the different 'versions', or at least sufficiently similar that a common approach is likely to be adopted. The next section will therefore consider the content of BPR as it is presented by a number of different authors to see if they share a common view of what BPR involves.

4 Content of BPR

There have been many attempts in the BPR literature to define the concept, not least by Hammer & Champy [29] who devote a whole chapter of their book to an elaboration of their statement that it is "the fundamental rethinking and radical redesign of business processes to achieve dramatic improvements in critical, contemporary measures of performance, such as cost, quality, service and speed". Unfortunately, few of the definitions show more than a limited consensus on its key features, particularly when the terms employed are examined in more detail. While some of the differences revolve around distinctions that would probably have appealed to mediaeval theologians, others mark more significant divisions within the field.

Process orientation

Perhaps the only real common ground between all BPR approaches is their view of organisations as being made up of *processes*. Hammer & Champy [29] define a process as "the collection of activities that takes one or more kinds of input and creates an output that is of value to the customer". This is contrasted with the traditional perspective which sees organisations as made up of separate functions (such as marketing, finance, sales and so on), typically located in their own specialist departments.

While agreed that BPR involves a focus on processes, there is disagreement in the literature about their scope.

There would appear to be four different "levels" at which these may operate. At the lowest level there are individual work tasks, such as customer inquiry handling, which combine various different, if simple, activities. At the second are work processes which may operate within a single functional department. At the third level the processes operate across departments within an organisation, and at the highest level they are inter-organisational. Each of these are, in principle, compatible with Hammer & Champy's definition, particularly if we take the currently popular view of organisations as markets, in which customers may be internal as well as external.

Harrington's concept of a business process which he defines as "all service processes and processes that support production processes" [31] would seem to operate at the first two levels, that is, within established departmental boundaries. Similarly, Morris & Brandon [51] define a process as "larger than a task ...[but] smaller than an area of business such as operations, human resources or shipping", while Hall *et al* [27] state that "a process can be as narrowly defined as a single activity in a single function". Davenport & Short [14] also talk of redesigning "interpersonal processes" which "involve tasks within or across small workgroups typically within a function or department". Hammer & Champy [29], however, suggest that processes to be reengineered typically operate across traditional functional/departmental boundaries, a view shared by Johannson *et al* [36] and Davenport [13] amongst others. BPR, Hammer & Champy argue, involves "breaking down the functional silos" and Davenport [13] states that "adopting process innovations inevitably entails cross-functional and cross-organizational change". Although these writers also allude to the possibility that processes may cross organisational boundaries, this is the particular focus of Venkatraman [64] who talks of "business reconfiguration" and Short & Venkatraman [58] who discuss "business network redesign".

Scale of change

Partly related to the scope of processes is the question of the scale of change, both in terms of extent of reorganisation and the performance improvements sought. Hammer & Champy [29] argue for a resolutely radical approach, describing re-engineering as a "new beginning", a "blank sheet of paper". As Hammer [28] puts it "the point is not to learn what happens to form 73B in its peregrinations through the company, but to understand the purpose of having form 73B in the first place". Thus BPR, Hammer & Champy argue, means "discarding existing processes and replacing them with entirely new ones" so as to achieve a "quantum leap" [in performance]. This view is shared by Johannson *et al* [36] who argue that organisations should be aiming not just for marketplace parity, but for "market dominance". This requires them to "break the china", "to challenge the very purposes, principles and assumptions on which their business is based". Davenport & Short [14] also talk of "fundamental reshaping" and Davenport [13] of "effecting creative and radical change to realize order-of-magnitude improvements". Similarly, Heygate [32] argues that managers have to be "immoderate in their ambitions to improve processes", while Hall *et al* [27] argue that "effective transformation ... requires a clean slate approach to process redesign".

Other writers, in contrast, appear to envisage a more modest, evolutionary activity. Harrington [31], for example, describes BPR as involving the "improvement of inefficient processes" and this is echoed in Fintech [23]. Glover, quoted in Warren [65], also argues that "radically changing the whole ethos of the business may not be necessary: there may be just one division which needs attention". Morris & Brandon [51] do not specifically discuss the scale of change, but the general absence of the revolutionary rhetoric characteristic of Hammer & Champy [29] suggests that their approach is also more incremental.

This point has recently shown signs of becoming a point of serious schism in the BPR literature. Thus as, Yates [70] reports it has lead to a parting of the ways between Michael Hammer and James Champy. Hammer, for his part, likens reengineering to an organisational "neutron bomb" and argues that "trying to do a little bit of reengineering is like trying to be a little bit pregnant". Champy, in contrast, argues that "there are good things that we don't want to destroy ... there is a sense of pragmatism in older managers that I don't think companies can afford to lose". Davenport has also criticised "the myth of the clean slate", arguing that "designing with a dirty slate will often yield a more implementable process" [15].

The means of achieving BPR

Another difference between BPR approaches relates to their ideas on how it should be achieved. Thus, while most writers specify a methodology for conducting BPR, Hammer [28] offers only a set of principles which the exercise should seek to pursue. Even in Hammer & Champy [29], there is no specific method, rather there is a discussion of outcomes, of the roles involved, case studies of the "experience of process redesign" in several companies and a discussion of the most common errors that lead to reengineering failure. Avoid these it is argued "and you almost can't help but get it right".

It would seem clear therefore that Hammer wishes to avoid a formulaic approach to BPR and sees it as a way of thinking and achieving radical organisational change rather than as a rigorous procedure to be followed. This may be contrasted with the recommendations of Morris & Brandon [51] and Booz-Allen & Hamilton [5] who argue that the use of a systematic methodology is essential for BPR success (even if, on closer examination, this often turns out to be a set of vague guidelines of limited practical utility). This view appears to be particularly prevalent in the information systems and industrial engineering literature on BPR (see for example, [26, 42, 66]). Klein [43] suggests that these positions represent two different schools of thought on the approach to BPR which he calls the "intuitives" and "methodologists" and identifies the particular consultancy companies associated with each.

The role of information technology

A potentially important area of disagreement between the different BPR approaches, as was noted in the introduction, is the contribution of IT. Two different roles for IT need to be distinguished here: firstly its use as a tool for the support of the BPR activity, and secondly its position as an axial principle around which the redesign of business processes should be planned. Davenport [13] calls these two roles "IT as implementer" and "IT as enabler" though there are some problems with this terminology as is discussed below. With respect to the first of these there would seem general agreement that BPR may be usefully supported by IT-based tools and some approaches explicitly recommend their use. For example, Morris & Brandon [51] give considerable attention to the use of simulation to test new process designs and identify it as a specific stage in their method, and Davenport & Short [14] discuss the use of CASE tools in drawing process models. Others discuss BPR activities which involve IT support. For example Sutherland [61] discusses the use of "reengineering laboratories" which make use of a variety of IT systems, while Heygate [32] suggests that "the role of IT in building individual's' skills may well prove to be the most valuable application of information technology so far discovered".

On the question of whether reengineered processes should be based on the use of IT, however, there would appear to be a significant divergence of opinion. A number of the key writers, such as Davenport & Short [14] for example, see IT as playing a central role in BPR, making the consideration of IT "levers" one of the five stages in their BPR method and describing as an exemplar a case of "IT-driven process redesign" at Xerox. Similarly, Guha *et al* [26] talk of "IT-induced reengineering". The concept has also been picked up by a number of IT/IS consultants who present it as a new approach to applying IT in organisations. For example Heygate & Brebach [33] explain how their "decision engineering, [IT-supported] organisational flattening, technology scan and IT cost estimation" techniques can support an "IT-driven business strategy", and Aikins [1] discusses how BPR can make use of knowledge-based systems. In some cases BPR is even associated with specific types of IT, usually workflow software and document image processing [57]. These writers would therefore seem to share Davenport & Short's view that IT and BPR "have a recursive relationship ... each is the key to thinking about the other" [14]. From this perspective IT needs to be considered in the early stages of redesign so that "awareness of IT capabilities can – and should – influence process design" [14].

Other writers, however, argue that "BPR and IT are certainly not synonymous" [61] and that "in theory, BPR does not have to involve IT at all" [23], rather it is "the common-sense practice of redesigning business processes before bringing in technology" [21]. Harrington [31], for example, does not mention the use of IT and confines discussion of the use of computers to a section on "automation". Morris & Brandon [51] also state that BPR "is not an IT topic", while Rowland [55] argues that "those who say that IT is a requirement of BPR have missed the point". This does not mean that these writers deny that IT may contribute to BPR, but rather that they believe that process design should precede consideration of IT options. For example Shillingford [57] states that "[o]ne attraction of BPR is that it makes IT subordinate to business objectives ... in most cases 80% of the value of BPR comes out of examining the way things are organised ... [o]nly 20% comes from IT".

An intermediate position appears to be that IT can "enable" new ways of carrying out processes and that redesign should therefore "be undertaken with a full knowledge of how technology can help" [5]. Hammer & Champy [29], for example, devote a chapter to illustrating that shared databases, expert systems, telecommunications networks, decision support tools and so on, "break the rules that limit how we conduct our work" and that IT is therefore a "part of any reengineering effort" whose importance it is "difficult to overstate". This argument, however, smacks rather of technological solutions in search of suitable applications rather than a convincing case that the use of IT is an essential feature of any reengineering activity as they suggest. It may be contrasted, for example, with the statement by Holtham [34] that "very few of the successful case studies of BPR have been initiated out of the new opportunities offered by IT .... the development of technologies such as workflow, groupware or document image processing have no *a priori* connection to the need to reengineer business processes". The concept of IT as an "enabler" may therefore be rather closer to the IT-led position taken by Davenport & Short [14] than might at first appear to be the case.

It is also possible, however, to exaggerate the differences between authors on the role of IT. Thus, as has been noted, those who see IT as secondary may also discuss how it can "make it possible for processes to be carried out in a new way" [23], while the proponents of IT-driven approaches will emphasise that BPR requires a "carefully-considered combination of both technical and human enablers" [13]. The vagueness of terms such as "enabler" means that it is almost impossible to tell whether there is a real difference of interpretation in either theory or practice between the writers, but it would seem reasonable from the context in which these terms are used that there is at least a difference of emphasis.

5 Discussion

Having considered the content of BPR, attention may now be turned to the three sets of questions identified in the introduction to this paper for which it was hoped to find answers. Is there only one type of BPR or are there several alternatives? What is the substantive content of BPR and how much of this is new? What factors may account for the interest shown in it? Each of these will now be considered in turn.

One BPR or many?

The analysis of the literature would seem to suggest that BPR is not a unitary concept (and it would have been perhaps more surprising to find that it was). However, if each approach is not to be classified as a separate version depending on the particular context in which it is applied, then the distinctive characteristics of BPR, or of certain recognised versions of it, need to be identified in order to develop an appropriate categorisation. This raises the problem of which approaches should be recognised as legitimate. While precedence might suggest that the approaches of Hammer [28] and Davenport & Short [14] should be seen as definitive this would seem unsustainable on practical grounds if nothing else. Thus, despite the efforts of CSC Index, the consultancy with which Hammer developed reengineering, to register "business reengineering" as a trademark, this cannot prevent other companies from selling approaches, which might be quite different from that envisaged by Hammer, under a slightly different name. Nor will it stop others from describing projects as reengineering, even if they bear little resemblance to Hammer's description. This is evident from the literature (see for example [2, 18]) where analyses of "BPR success" refer to widely varying initiatives, many of which did not even describe themselves as reengineering.

A second problem with categorising approaches concerns how closely they need to correspond to an archetype to be classified as belonging to a particular version (or even to be excluded from the concept altogether) and how this might be assessed. For example two approaches may share a strong emphasis on the role of IT (assuming there is some reliable measure of strength of emphasis), but differ on the importance of a definite methodology (which again may be very difficult to define). It would therefore seem difficult to provide a robust basis on which to classify different forms of BPR.

An alternative perspective is put forward by Davenport [13] who suggests that, rather than discrete "schools" of BPR, there is a continuum of approaches between two extremes of "process improvement", an incremental, continuous, bottom-up approach, and "process innovation" which is radical, IT-led, one-off and top-down. Talwar [62] suggests a similar spectrum, from "process improvement" though "process reengineering", "business reengineering" and "transformation" to "ongoing renewal". While this takes account of the observed differences between approaches, it assumes that they vary consistently along all dimensions. In practice, this would not appear to be the case. For example, Johannson *et al* [36] might appear to be a good example of Process Innovation with their emphasis on radical organisational change. However they clearly view their approach as originating with TQM, and see IT as only one among a number of transformatory factors.

It would seem therefore that it is likely to be impossible to agree a precise definition of BPR that does not exclude some of the approaches that have been associated with the concept, or to identify a limited number of distinct versions. It may thus be best to regard BPR simply as an umbrella term for a set of process-oriented approaches to significant organisational change.

What's new?

The limitations of such a broad, inclusive definition of BPR are obvious, however, when considering the possible novelty of the concept. For example, most of the literature acknowledges that process thinking has been at the heart of the "quality revolution" and Morris & Brandon [51] also state that the techniques on which they claim BPR is based (IE, time and motion, Operational Research and Systems analysis) "have all been concerned with processes for several decades".

If the core idea on which BPR is founded is shared with many other techniques, then what is the basis for claiming that it offers something significantly different from the others? What is to say that it is not simply a well-marketed repackaging of them? One of the key themes of a substantial proportion of the BPR literature which might therefore be a novel feature, is the contribution of IT. Certainly, the BPR literature typically lays considerably more emphasis on IT than a number of the concepts, such as TQM and "Excellence", that immediately preceded it.

The idea of IT as a "driver" of organisational change, however, goes back at least to Leavitt & Whisler [47]. Moreover, even if it is claimed that recent technological developments have made existing uses of IT more effective and provided opportunities for new and more significant interventions, the precise role that IT should play is a source of considerable disagreement amongst writers on the subject. It would thus seem an unsatisfactory basis for defining the unique character of BPR, particularly as many of its proponents are seeking to move away from the sort of technocentric viewpoint that such a definition would imply.

The remaining distinctive feature of BPR would therefore seem to be the scale of the change involved. As has been noted, however, there are some authors who do not seem to share the enthusiasm of Hammer & Champy [29] for "fundamental rethinking and radical redesign". Even if it were possible to exclude these faint-hearts from associating their approaches with BPR by some mechanism, though, this would still not resolve the problem of how the scale of change is to be measured and how the boundary between BPR and the other process-oriented techniques could be defined. For example, does BPR only describe approaches which seek change across organisational boundaries? If so, how many boundaries need to be crossed? All? More than one? Should change be measured by intent or by outcome? If the former, then BPR risks being discredited by approaches which promise more than they can deliver. If the latter, then a way is needed of separating out the contribution of the BPR activity to the outcome, from those due to other changes which may have been going on at the same time. For a variety of reasons therefore, it would appear that scale of change also does not provide a reliable means of defining BPR.

The final way in which BPR might be argued to be novel is its particular combination of process thinking, creative use of IT and radical organisational change. While there may be some practical value to this viewpoint in terms of enabling BPR writers to differentiate their approach from other techniques by emphasising some particular aspect of this combination as it suits their case, the analysis of the literature would seem to suggest that the degree of variation on each of these dimensions is so great that 'novelty' of BPR would vary from approach to approach.

This would seem to suggest, therefore, that despite BPR being widely hailed as a distinctive new contribution to management thinking (the endorsement of Hammer & Champy's book by no less an authority than Peter Drucker would suggest that this this view is not confined to just a few cranks), it seems very difficult to identify clearly what it is that makes it so significant. In large part, this problem stems from Hammer's definition of BPR in terms of practice. In the absence of any explicit, new theoretical principle on which the concept may be said to be based, Hammer requires BPR to be judged by the actions it gives rise to – if it achieves new forms of organisational behaviour which significantly improve organisational performance in some agreed way, then it is a significant new technique. As has been noted, however, this does not provide the basis for an unambiguous definition of the term. Moreover, as Grint [25] points out, all of the changes that Hammer & Champy [29] identify as characterising the "new world of work" in the reengineered organisation have been proposed by earlier writers. Thus even the changes in organisational behaviour associated with BPR may not be new either.

In the traditional interpretation of Kuhn's theory of scientific revolutions [44], therefore, it would seem difficult to argue that BPR constitutes a new paradigm, since it does not offer a new perspective on organisations which differs radically from TQM and other improvement programmes that preceded it. In Lakatos' terms [45], both appear to be operating within the same Scientific Research Programme. It could hardly be claimed, moreover, that TQM, for example, has not been extremely influential in changing management thinking and that BPR was thus in some way a conceptual Copernicus building on the unrecognised contribution of TQM's Tycho Brahe. Yet, if it is accepted that the immense interest shown in BPR is based on something more than self-delusion, then we need to provide some explanation of what this particular Emperor is wearing, even if it does not seem possible to define this in absolute terms.

To do so, however, would seem to require a reassessment of the nature of imperial regalia, to argue that the Emperor's old clothes were equally insubstantial. TQM, Taylorism or any of the other techniques which we may wish to consider as underlying earlier forms of organisational practice are/were, just as much as BPR, based on a set of socially-sustained beliefs which may or may not have had 'real' substance, but which depended on continuing faith in their efficacy. This is not simply a sleight of hand, but is based on the analysis of science and technology developed by Callon [9], Latour [46] and others which suggests that even in the so-called "hard" sciences, knowledge is not "discovered" in an external "reality", but is constructed through the recruitment of a network of allies who underwrite a particular viewpoint. Paradigm shifts therefore result from the ability of concepts to recruit the most influential allies. In this, rhetorical power may be rather more important than the "truth" of their views. The final question (why is BPR so popular?) may therefore help us to understand the allies that Hammer and Davenport & Short have, wittingly or otherwise, recruited and hence how BPR has come to have such an impact on current management thinking.

Why is BPR so popular?

The answer given by Hammer & Champy is that corporate America (*sic*) is facing a crisis, which they characterise in terms of three Cs – customers (who are taking charge), competition (which is intensifying) and change (which is becoming constant). Solutions are needed to enable companies to succeed in this new world of business. BPR, they argue, is the only technique which is capable of providing one.

Whether or not it is accepted that this crisis is real and enduring, for example Wood [69] questions similar claims made in relation to the flexibility debate, the concept that the business environment has recently become significantly more unstable and that substantial organisational change is unavoidable, appears to be widespread, particularly in the US, at present. BPR therefore appears to capture the spirit of the times in proposing that a totally new approach is needed. Such a viewpoint may also be seen as serving to legitimate the adoption of radical measures. If the company cannot hope to survive without complete re-organisation and the shedding of "armies of unproductive workers" [28] then these become a necessary price to pay. Thus, as Dixon *et al* [18] comment, "a crisis – real or perceived – may be necessary to create the conditions required to attempt a reengineering effort".

As Hammer & Champy observe [29], though, they are not the first to have diagnosed a crisis or to have suggested the way out of it. Many other techniques, from "management by objectives ... [to] ... one-minute managing" have been put forward in the past. None of them, however, "has reversed the continuing deterioration of America's corporate competitive performance". This is taken by Hammer & Champy as evidence that they were passing fads which "have only distracted managers from the real task at hand". The failure of these techniques is therefore used to bolster the case for BPR.

Paradoxically, these earlier techniques also help to sustain BPR. Thus, for example, the notion that new solutions have to be continuously invented provides a reason for looking to reengineering. The precedents for process-based techniques, even if they have not been particularly successful, also provide legitimation for the reengineering approach. In particular, the continuing influence of TQM, despite Hammer & Champy's dismissal of it, means that process-thinking may be seen as a distinct element in the network.

Hammer & Champy's particular focus on the problems of US companies may be seen as the recruitment of US patriotism as another ally. Thus they write that reengineering is "not another idea imported from Japan", but "capitalizes on the same characteristiscs that have traditionally made Americans such great innovators". Rather than "try to change the behaviour of American workers and managers" it "takes advantage of American talents and unleashes American ingenuity".

The success stories are another important element of the BPR network, serving both to affirm the potential of the technique, but also to associate it with the named corporations. Thus the 'hard facts' of the performance improvements cited by Hammer & Champy provide quantitative support for their claims which are thereby reinforced against sceptical questioning. Moreover the 'fact' that Ford, Bell Atlantic or Hallmark cards have successfully adopted reengineering provides powerful evidence of its value. Other case-study companies may be less well-known, but their inclusion shows that it is effective for all organisations not just those in the premier league.

Success stories also relate not just to the effectiveness of the technique, but to those who sell it. For example the near sixfold increase in the revenues of CSC Index in the last five years, is said to have "propelled [it] ... to the forefront of the [US management] consulting business" [8]. IDC [35] estimated a world market for BPR of $230 million in 1993 which they suggested was growing at 46% per annum.

These sorts of claims and figures draw in a variety of human actors to the reengineering network. Clients would seem likely to be attracted by the chance of order of magnitude performance improvements. As more leading companies adopt BPR, the need for others to join the bandwagon would seem likely to increase, both to show themselves to be up with the latest developments and to ensure that potential benefits are not lost.

For executives of the many companies which, in Hammer & Champy's words are "bloated, clumsy, rigid, sluggish, non-competitive, uncreative, inefficient, disdainful of their customer needs and losing money", the message that BPR provides the way out of this situation would seem a very seductive one. Hammer & Champy's insistence on the need for charismatic leadership of a firmly top-down process may also be expected to be attractive to those whose position this reinforces or who aspire to such roles. Similarly, those working in IS and/or Industrial Engineering may be expected to be attracted by the prominence given to their role in many versions of BPR.

For management consultancies looking for new opportunities, the size and relatively under- developed state of the BPR market (giving low entry barriers, few established players and significant scope for product differentiation) would seem to have considerable potential. The investment being made by the major consultancies in developing BPR products illustrates their recognition of this potential [63], but also serves as a signal of reengineering's importance to competitors and clients.

For IT consultancies BPR provides both an opportunity and a threat. If a suitable "front-end" product can be developed then there may be opportunities to capture market share in a new and rapidly-expanding area. Those who fail to offer BPR services, however, may find that part of their market has been sequestered both by traditional competitors who have taken the initiative and by new entrants from the management consultancy field. By getting in early on the BPR act, IT consultancies may also be able to strengthen the links between IT and BPR and thereby increase the size of their market.

The effectiveness of BPR as a management concept may also be due not just to the claims it makes, but to the way it makes them. As Jones [37] discusses, the violent language adopted by Michael Hammer, with its talk of "don't automate, obliterate", neutron bombs and grinding and frying out organisational fat gives the concept a strongly macho flavour. The apocalyptic rhetoric of organisational redemption through destruction of old habits and the adoption of pure, new processes also has biblical overtones. This is echoed in the comments of Stewart [59] who describes Hammer as "reengineering's John the Baptist, a tub-thumping preacher" and by Hammer himself who talked of reengineering as a "theology because it requires a belief that there is a new way of doing things ... it requires faith" (quoted in [7]).

Finally, it may be argued that process thinking itself simply has a considerable intuitive and practical appeal. Whether positive associations with the Quality movement are made explicitly or not, for individuals accustomed to a traditional functional perspective on organisations it can be seen to constitute a good example of an "interesting proposition" in terms of the classification scheme of Davis [16] . Thus it suggests that what had traditionally seemed to be a phenomenon that functioned effectively as a means to an end, ie the division of labour along functional lines to encourage the development of specialist expertise would ensure maximum organisational performance, was in reality a phenomenon that functions ineffectively.

6 Fad, Failure or Fatally Flawed – The Future of BPR?

This diverse network of allies recruited by BPR may help to account for its current position as "the hottest trend in management" [60]. Not surprisingly perhaps, it has also attracted many detractors. Thus Talwar [62] describes BPR "as the most fashionable and potentially the most detested management concept of the 90s". Many of these critics seek to show that BPR is simply a passing fad whose fall will be almost as meteoric as its rise. Holtham [34], for example, provides a timetable for the BPR lifecycle which predicts its demise in 1996-98. Those who dismiss BPR in this way generally adopt two lines of attack: arguing either that it is nothing more than a rebadging of existing concepts, or that its success owes more to skilful packaging than to any substantially new contribution to management thinking.

As has been discussed, the case for doubting BPR's novelty is fairly strong. The particular reasons offered by different authors for arguing that "we have seen it all before", however, often provide more insight on their own background than on the theoretical antecedents of BPR. Thus IS people see BPR as a reworking of systems analysis, industrial engineers as an extension of O&M and quality experts as a rebadging of TQM concepts. There can also sometimes appear to be an element of defensiveness in such arguments that suggests that the authors see this upstart concept as a threat to their expertise (and lucrative consultancy opportunities?).

Criticisms of the hype surrounding BPR also have some validity. As many authors note, the concept has clearly become one of the key management buzzwords of the 1990s and there is plenty of evidence of a bandwagon effect in operation. Stewart [60] even reports one telephone company executive as stating that "if you want to get something funded around here – anything, even a new chair for your office – call it reengineering on your request for expenditure". Such attitudes, however, do not necessarily prove that BPR is not a significant development. Thus, while it is possible to criticise the particular rhetorical style adopted by Michael Hammer, BPR is certainly not unique amongst management concepts in being promoted in such robust and evangelical terms. Indeed as Eccles & Nohria [20] have argued, effective use of rhetoric is an important aspect of successful change management. Moreover, as Hammer & Champy [29] are keen to argue, the proof of the reengineering pudding is in the eating.

This may help to explain why Hammer & Stanton [30] so vigorously contest what they see as the mythology of reengineering failure, arguing that the widely-cited figure of 70% was made up by Michael Hammer, based on experience with early reengineering projects. Now that the concept is much better understood, it is suggested, there is "little excuse" for companies to fail. If they do, it is because they fail to apply it correctly. This view is implicitly supported by a thriving sub-genre of the BPR literature which describes how to ensure reengineering success (see for example [4, 27]). Unfortunately, the empirical support for these recommendations, as with much of the writing on BPR, is generally weak and their prescriptions often amount to little more than 'be a successful company and do it right'.

Moreover, even if they do not fail completely, BPR initiatives may deliver less than is promised by their proponents. Thus Moad [49] reports that the benefits obtained from BPR projects intended to increase revenue were rated as achieving only 2 on a scale from 1 to 10, and even the best scored just 5 out of 10. James Champy is also reported as conceding that 30–40% of BPR initiatives "disappoint" [48]. A further interesting slant on this question is provided by Craig & Yetton [12] who point out that the rewards from an incremental improvement programme over a number of years may exceed those of a one-off radical change programme (and with less risk).

The claim made by Hammer & Stanton [30] that "the fault [for reengineering failure] lies not in reengineering, but in ourselves" may also be challenged on more substantial grounds. Thus, as Jones [38] discusses the BPR concept may be seen to have many internal contradictions. One of these is neatly illustrated by King [41] who notes that "the most problematic aspect of reengineering is that people are often asked to be creative so that their jobs (or those of their colleagues) can be eliminated or drastically changed". Similarly, Wheatley (quoted in [6]) argues that reengineering will inevitably fail as its top-down, mechanistic approach stifles learning and creativity.

Does this mean therefore that BPR is past its sell-by date and that forward-looking organisations should seek elsewhere for their inspiration? Eccles & Nohria [20] argue that the perceived failure of established management concepts is a necessary precursor of the emergence of new "solutions". The "open season" for BPR critics noted by Hammer & Stanton [30] may therefore be an ominous sign for the concept's longevity. The on-line search also indicates that there has been a significant decline in the number of BPR-related articles since mid-1994. However, despite a number of papers discussing what is "beyond" reengineering [15] or "after" it [50], the absence, as of early 1995, of any obvious successor that is capturing the management imagination, suggests that reports of its imminent demise may be premature. Moreover, unless organisations are to be caught up in a continuous cycle of change for change's sake, and are to learn nothing from past experience, then it is important that the valuable elements of BPR are not lost, even if the label itself becomes out-dated. It is hoped that the analysis provided in this paper, may help in identifying these elements.

References

1. J Aikins. 'Business process reengineering: where do knowledge-based systems fit'. *IEEE Expert*, **8**(3), (1993).
2. A Ascari, M Rock, and S Dutta. 'Reengineering and organizational change: lessons from a comparative analysis of company experiences'. *European Management Journal*, **13**(1), (1995).
3. P Bartram. *Business Reengineering:The Use of Process Redesign and IT in Transforming Corporate Performance*. Business Intelligence, London, 1992.
4. B J Bashein, M L Markus, and P Riley. 'Preconditions for BPR success: and how to prevent failure'. *Information Systems Management*,, **11**(2), (1994).
5. Booz-Allen and Hamilton. Business process redesign: an owner's guide, 1993.
6. T Brown. 'De-engineering the corporation'. *Industry Week*, (18 April 1994).
7. J A Byrne. 'Management's new gurus'. *Business Week*, (31 August 1992).
8. J A Byrne. 'Reengineering: beyond the buzzword'. *Business Week*, (24 May 1993).
9. M Callon, J Law, and A Rip, eds. *Mapping the dynam-*

ics of science and technology: sociology of science in the real world. Macmillan, Basingstoke, 1986.
10. A Cane. 'No guarantee of success - business process reengineering'. *Financial Times,* (26 September 1994).
11. C J Coulson-Thomas. 'Corporate transformation and business process reengineering'. *Executive Development,* **6**(1), (1993).
12. J F Craig and P W Yetton. 'The dual strategic and change role of IT: A critique of business process reengineering'. Working Paper 94-002, Australian Graduate School of Management, Kensington, NSW, (1994).
13. T H Davenport. *Process Innovation: Reengineering Work Through Information Technology.* Harvard Business School Press, Boston, Mass, 1993.
14. T H Davenport and J E Short. 'The new industrial engineering: information technology and business process redesign'. *Sloan Management Review,* (Summer 1990).
15. W H Davidson. 'Beyond re-engineering: the three phases of business transformation'. *IBM Systems Journal,* **32**:65–79, (1993).
16. M Davis. 'That's interesting! towards a phenomenology of sociology and a sociology of phenomenology'. *Philosophy of the Social Sciences,* **1**, (1971).
17. S F Dichter. 'The organization of the 90s'. *McKinsey Quarterly,* **1**, (1991).
18. J R Dixon, P Arnold, J Heineke, J S Kim, and P Mulligan. 'Business process reengineering: improving in new strategic directions'. *California Management Review,* **36**(4), (1994).
19. M Earl. 'The new and old of business process redesign'. *Journal of Strategic Information Systems,* **3**(1), (1994).
20. R G Eccles and N Nohria. *Beyond the Hype: Rediscovering the Essence of Management.* Harvard Business School Press, Boston, 1992.
21. Fintech. 'Business process redesign: IT suppliers take lesons from users'. *FinTech Electronic Office,* (9 September 1992).
22. Fintech. 'Rank Xerox to sell business process redesign consultancy and tools'. *FinTech Electronic Office,* (6 May 1992).
23. Fintech. 'Spotlight: Business process redesign brings big benefits'. *FinTech Electronic Office,* (2 December 1992).
24. C Grey and M N. 'Re-engineering organizations: a critical appraisal'. *Personnel Review,* **24**(1), (1995).
25. K Grint. 'Reengineering history: social resonances and business process reengineering'. *Organization,* **1**(1), (1995).
26. S Guha, W J Kettinger, and J T C Teng. 'Business process reengineering: building a comprehensive methodology'. *Information Systems Management,* (Summer 1993).
27. G Hall, J Rosenthal, and J Wade. 'How to make reengineering really work'. *Harvard Business Review,* **71**, (Nov-Dec 1993).
28. M Hammer. 'Reengineering work: don't automate, obliterate'. *Harvard Business Review,* **68**, (Jul-Aug 1990).
29. M Hammer and J Champy. *Reengineering the Corporation: a Manifesto for Business Revolution.* Nicholas Brealey, London, 1993.
30. M Hammer and S Stanton. 'No need for excuses'. *Financial Times,* (5 October 1994).
31. H J Harrington. *Business Process Improvement.* McGraw Hill, London, 1991.
32. R Heygate. 'Immoderate redesign'. *McKinsey Quarterly,* **1**, (1993).
33. R Heygate and G Brebach. 'Corporate reengineering'. *McKinsey Quarterly,* **2**, (1991).
34. C Holtham. 'Business process re-engineering: contrasting what it is with what it is not'. In *Business Process Re-engineering: Myth and Reality.* Kogan Page, London, (1994).
35. IDC. Business process redesign: the confluence of management and information technology consulting, 1992.
36. H J Johannson, P McHugh, A J Pendlebury, and W A Wheeler. *Business Process Reengineering: Breakpoint Strategies for Market Dominance.* John Wiley, Chichester, 1993.
37. M R Jones. 'Don't emancipate, exaggerate: rhetoric,reality and reengineering'. In R Baskerville, S Smithson, O Ngwenyama, and J I DeGross, eds., *Transforming Organizations with Information Technology,* Amsterdam, (1994). North Holland.
38. M R Jones. 'The contradictions of reengineering'. In G Burke and J Peppard, eds., *Examining Business Process Reengineering.* Kogan Page, London, (1995).
39. J M Juran. *Breakthrough Management.* Macmillan, London, 1964.
40. R B Kaplan and L Murdock. 'Core process redesign'. *McKinsey Quarterly,* **2**, (1991).
41. W R King. 'Process reengineering: the strategic dimension'. *Information Systems Management,* **11**(2), (1994).
42. M M Klein. 'IEs fill facilitator role in benchmarking operations to improve performance'. *Industrial Engineering,* (September 1993).
43. M M Klein. 'Reengineering methodologies and tools'. *Information Systems Management,* **11**(2), (1993).
44. T S Kuhn. *The Structure of Scientific Revolutions.* Chicago University Press, Chicago, 1961.
45. I Lakatos. 'Falsification and the methodology of scientific research programmes'. In I Lakatos and A Musgrave, eds., *Criticism and the Growth of Knowledge.* Cambridge University Press, Cambridge, (1970).
46. B Latour. *Science in Action.* Open University Press, Milton Keynes, 1987.
47. H J Leavitt and T L Whisler. 'Management in the 1980's'. *Harvard Business Review,* **36**, (Nov-Dec 1958).
48. C Lorenz. 'Putting reengineering in perspective'. *Fi-*

nancial Times, (21 October 1994).
49. J Moad. 'Does reengineering really work?'. *Datamation*, (1 August 1993).
50. J Moad. 'After reengineering: taking care of business'. *Datamation*, (15 October 1994).
51. D Morris and J Brandon. *Reengineering Your Business*. McGraw Hill, London, 1993.
52. E Mumford. 'New treatments or old remedies: is business process redesign really socio-technical design?'. *Journal of Strategic Information Systems*, **3**(4), (1994).
53. F Ostroff and D Smith. 'The horizontal organization'. *McKinsey Quarterly*, **1**, (1992).
54. J Parker. 'An ABC guide to business process reengineering'. *Industrial Engineering*, (May 1993).
55. P Rowland. 'The process for improving business'. *The Independent*, (22 June 1994).
56. M Scott-Morton, ed. *The Corporation of the 1990s - IT and Organizational Transformation*. Oxford University Press, Oxford, 1991.
57. J Shillingford. 'Offices follow the factory's example'. *Financial Times*, (19 March 1992).
58. J E Short and N Venkatraman. 'Beyond business process redesign: redefining baxter's business network'. *Sloan Management Review*, (Fall 1992).
59. T A Stewart. 'The search for the organisation of tomorrow'. *Fortune*, **125**(10), (1992).
60. T A Stewart. 'Reengineering: the hot new management tool'. *Fortune*, **128**(4), (1993).
61. D Sutherland. 'The reengineering lab'. *Insights Quarterly*, **4**(3), (1992).
62. R Talwar. 'Re-engineering – a wonder drug for the 90s'. In C Coulson-Thomas, ed., *Business Process Re-engineering: Myth and Reality*. Kogan Page, London, (1994).
63. J Thackray. 'Fads, fixes and fictions'. *Management Today*, (June 1993).
64. N Venkatraman. 'IT-induced business reconfiguration'. In M Scott-Morton, ed., *The Corporation of the 1990s - IT and Organizational Transformation*. Oxford University Press, (1991).
65. E Warren. 'Streamlining from the outside in'. *Informatics*, (May 1993).
66. D G Wastell, P White, and P Kawalek. 'A methodology for business process redesign: experience and issues'. *Journal of Strategic Information Systems*, **3**(1), (1994).
67. H Willmott. 'Business process reengineering and human resource management'. *Personnel Review*, **23**(3), (1994).
68. J P Womack, D T Jones, and D Roos. *The Machine That Changed the World*. Rawson Associates, New York, 1990.
69. S Wood. 'The transformation of work?'. In S Wood, ed., *The transformation of work?* Unwin Hyman, London, (1989).
70. R E Yates. 'Two separate tracks for re-engineering's engineers'. *Chicago Tribune*, (July 24 1994).

Received: June 1995

Editorial

What is a good contribution?

The work of refereeing is a thankless task that infringes on the limited time of those who are best in the field. Nevertheless, it is important since the quality of refereeing does significantly impact on the quality of the final product. In order to be fair to those limited and committed referees involved, I am of the opinion that we need to become more discerning in what we put before them. There is, of course, the important role of a journal (especially so with SACJ) to assist young researchers to develop their skills by getting feedback from competent reviewers in the field. However, one must not "abuse" this privilege by sending anything "just in case"—it is the academic responsibility of the authors to ensure that the paper submitted is of high standard.

What, then, is a good contribution? A good contribution has essentially three dimensions [1]: relevance, rigor and impact—of which relevance is the most important. Although these dimensions were formulated in the context of a discussion on Information Systems research, they may also have relevance for Computer Scientists. To assist potential authors in assessing their contributions (in the making), here are some questions, steps or criteria that should clearly and explicitly be addressed before submitting a paper to SACJ.

Relevance

- Who is the target audience of influence (IS mangers, system designers, educators, etc.)?
- What is the concern in the target audience that the paper addresses?
- Why is this a concern in the target audience?

Rigor

- What is the wider intellectual context of the study? Place it in this wider context.
- What other disciplines (or fields within the discipline) have studied or are concerned with the problem in the paper?
- What is the best research methodology for this problem domain?
- Why is the proposed research methodology selected?
- What question/problems may a referee raise and how are they being addressed?

Impact

- What contribution is the paper making? Is this clear and explicit in the conclusions?
- Why would a practitioner want to read the paper?
- What would a top scholar in the field say about the paper?
- Would a supervisor recommend the paper to his student(s)?
- Why can the journal not afford to publish the paper?

Clearly all of these questions may not always be equally relevant, and there may be others that could be relevant in assessing the contribution of a paper. Nevertheless, potential authors should rigorously critique their contributions and not merely "dump" material for publication. We owe it to ourselves and to the community at large.

News

Readers and contributors will be interested to hear of various plans for SACJ. These have evolved during a series of meetings held between editorial staff and available members of the editorial board and are briefly outlined below.

- Much time has been spent discussing whether or not SACJ should have a particular and/or exclusive focus. Should it seek a niche market, such as only publishing articles dealing with IT and development? Our deliberations have lead us to reject such an exclusive focus. Instead, SACJ will build on its traditional role of being primarily, but not exclusively, a platform for IT researchers in South Africa. In addition, the editorial staff intends widening horizons by actively recruiting contributions, primarily from other Southern African countries, but also from countries elsewhere in Africa. While SACJ will thus continue publishing a miscellany of articles that reflect the local status of IT research, it hopes to be a vehicle that encourages, stimulates and eventually reflects research on the African continent as a whole.
- SACJ specifically encourages contributions that synthesise existing research results, such as survey articles, reviews and taxonomies. Where such articles are of significant scope, they are important for research. They will therefore be placed in the research section of the journal and may, in consequence, be submitted to the Department of National Education (DNE) for subsidy purposes.
- In order to sharply differentiate the research section from the non-research section of SACJ, page numbers in the Communications and Reports section will henceforth be preceded by an A. This is intended to signal that articles in the Communications and Reports section should not be submitted to the DNE for subsidy purposes.
- There have been strong pleas that the journal be brought out on a more regular basis. To date, issues have been held back until a fairly sizeable number of articles are available for publication. This has resulted in two to

three issues per year. In future we intend bringing out four issues, two of which will be regular issues, and the other two will be special issues dealing with some topical theme. This process has already started in that the next issue (in October) will be on IT and Development. A further special issue on Networking/Telecommunications is planned for mid-1996. Articles for the former special issue are currently being reviewed. Potential contributors to the latter special issue might want to start thinking about possible publications, but should delay submission of manuscripts until further details are provided in a formal call for papers.

- To date, SACJ has not obliged contributors to sign away copyright. Although the matter of copyright has become somewhat obscured in the age of electronic documents, the Internet, WWW, etc., SACJ will in future require that contributors sign a copyright form before their work is published. In dealing with issues relating to electronically available material, we shall follow guidelines proposed in recent issues of the Communications of the ACM.
- Some readers may have noticed that SACJ now appears on the World-Wide Web. This is due to the efforts of the production editor. The WWW home page address is provided on the front inside cover of the journal. Extensive information to potential subscribers and contributors has been provided. This includes information on submission procedures, document preparation, etc. Abstracts of published articles are also given.
- Presently SACJ is sent to Inspec where selected articles are indexed. Arrangements are currently being made for Science Citation Index to provide a similar service. This should significantly increase exposure of SACJ's contents to international researchers.
- SACJ encourages contributions from young and/or inexperienced researchers and will assist them in various ways to attain the quality required by the journal. For example, they may be provided with additional editorial assistance (perhaps to be paid for, if warranted by the scope of the assistance). Referees will be urged to be encouraging in their feedback and to formulate feedback clearly and specifically in ways that will facilitate the revision process. Potential contributors may also request to be put in touch with a mentor to advise in planning and writing up research. As a further encouragement to young researchers, consideration is being given to an annual citation for the best student contribution.

Derrick Kourie and Lucas Introna
Editors

1. Keen, Peter,G.W. 'Relevance and Rigor in Information Systems Research: Improving Quality, Confidence, Cohesion and Impact' in Nissen *et al*, eds., *Information Systems Research: Contemporary Approaches and Emergent Traditions*. IFIP 8.2, 1991.

SACJ is produced with kind support from
Mosaic Software (Pty) Ltd.

Training Strategies for Architecture-specific Recurrent Neural Networks

J Ludik I Cloete
Department of Computer Science, University of Stellenbosch, Stellenbosch, 7600, South Africa

Abstract

The typical approach taken by other researchers to address the defects of standard backpropagation is to investigate alternative methods for selecting initial parameter values, or for adjusting parameter values during BP training. We have suggested that the training strategy, *that is the method for presentation of examples to the network during learning according to some performance criteria, is a viable alternative method to produce an effective solution within a feasible time. The dual purpose was to evaluate training strategies and* architecture-specific recurrent neural networks *(ASRNNs) simultaneously for different applications, varying in complexity and ranging from sequence recognition to sequence generation. It was demonstrated with six different ASRNNs and feedforward networks that for several applications,* Increased Complexity Training *(ICT) outperformed* Combined Subset Training *(CST) and* Fixed Set Training *(FST). We have also compared several ASRNNs for the Counting and Addition tasks, and found the Output-to-Hidden Hidden-to-Hidden ASRNN, one of our proposed ASRNNs, to be the most superior architecture, outperforming the conventional ASRNNs between 44% and 87% for the training strategies. For both ICT and CST the method of determining the required RMS termination criteria per subset was unsatisfactory since it required experimentation in stead of being performed algorithmically. In answer to the latter need, we have proposed incremental training strategies,* Incremental Subset Training *(IST) and* Incremental Increased Complexity Training *(IICT), which improved the convergence rate compared to CST, FST, and even ICT. We have also proposed six Delta training strategies by first employing the Delta Ranking Method, which determine the complexity relation between the input patterns by obtaining their inter-pattern distances and then ranking them according to some scheme. We have introduced three basic ranking schemes which led to* Smallest Delta Subset Training *(SDST),* Largest Delta Subset Training *(LDST),* Alternating Delta Subset Training *(ADST), their incremental versions, and also their epoch versions. All the Delta training strategies proved to be very effective (evaluated with different applications) in reducing the training time when compared to the conventional strategies. Incremental Delta training strategies performed the best overall, signalling that the ordering of training patterns according to our proposed delta ranking schemes, especially when presented in an incremental fashion, forces the network to discriminate between classes early in the training process, leading to reduced training time. The training strategies should be regarded as different tools in a training strategy toolbox, each one suited for its particular purpose.*

Keywords: *Training Strategies, Architecture-specific Recurrent Neural Networks, Incremental, Increased Complexity*
Computing Review Categories: *I.2.6, I.2.8*

1 Introduction

Much attention has been given to improving the learning time of neural networks trained with the popular back propagation (BP) algorithm. The reason is that because BP is a variation of Steepest Descent the inherent ill-condition of the algorithm makes it slow or prone to non-convergence. Previous investigations considered the initial values and changes to various parameters and variables of a neural network during training, such as the learning rate, momentum and weights [11], to speed up learning. However, training strategies have mostly been overlooked as a possible method to address the problems of BP. For *architecture-specific recurrent neural networks* (ASRNNs), also called Simple Recurrent Networks, these deficiencies of BP are more pronounced. These networks have feedback connections that are organized in strict architectures such as hidden-to-hidden, output-to-hidden, and output-to-output unit feedback. The purpose of this paper is firstly to develop training strategies for ASRNNs and feedforward networks that do not only alleviate these problems, but also lead to considerable improvement in training time. The second goal is to evaluate different ASRNNs with the training strategies by comparing the training performance of Elman, Jordan, and Temporal Autoassociation ASRNNs as well as other newly constructed ASRNNs. Besides the problems with BP, the third goal is to address problems such as the effect of multiple attractor basins, the size of the initial subset, the increment in the subset size, the termination criteria for each subsequent subset, and the number of examples required for good generalization. The final goal is to examine the training strategies and ASRNNs with different applications, varying in complexity and ranging from sequence recognition to sequence generation.

In this paper we have developed *Increased Complexity Training*, two *incremental* training strategies and six *delta* training strategies, and showed that they improve training performance significantly and abate the BP problems for different ASRNNs as well as feedforward networks. We have also used the training strategies to compare different ASRNNs and found one of our suggested architectures, the Output-to-Hidden Hidden-to-Hidden ASRNN, to be the most efficient architecture. The incremental training strategies also effectively address the problems mentioned in goal three. The training strategies and ASRNNs are evaluated with tasks involving only input sequences (Counting), input and output sequences (Addition), and sequence generation (Digital Morse code generation). The feedforward networks are examined using a Cluster detection and Digit recognition task.

2 A Taxonomy of Training Strategies

For orientation purposes, a taxonomy of existing and newly suggested training strategies is presented. An example of an existing training strategy is Cottrell & Tsung (1991)'s *Combined Subset Training* (CST). CST first entails selecting randomly a manageable subset of training patterns to train the network initially. Then when the network has learned this set fairly well, select randomly another subset of equal size and add it to the first. Train the network further with the combined set. Repeat this procedure until the whole training set is included, or until the network is able to generalize to the rest of the original training set. In this chapter we introduce various training strategies and illustrate experimentally that with all other network parameters identical, the training strategy alone can speed up learning time considerably.

In the conventional *Fixed Set Training* (FST) the same set of training patterns are repeatedly presented during BP training until a success criterion is met. Our proposed training strategies investigate several points for variation of the training schedule: (1) construction of training subsets of the fixed set; (2) ordering of subsets according to a particular criterion; (3) termination criteria for each training subset; and (4) when weight updates occur. We briefly discuss these issues in turn, before elaborating on them in subsequent sections.

(1 & 2) Construction and ordering of subsets: There are many possible ways to construct and order subsets of the fixed training set. We have firstly proposed *Increased Complexity Training* (ICT) [1] in which the fixed set was partitioned into subsets and ranked, each subset being more complex than the previous. The complexity measure was problem-dependent, e.g. the number of output units switched on by patterns in a subset for a counting problem, or subsets of one, two and three column examples for an addition problem. The network is first trained on the least complex subset. Then the second subset is randomly merged with the first, and the network trained again until successful. This process is repeated until the entire fixed set has been used. To circumvent the problem-dependent complexity measure, we have also proposed *Delta Subset Training* (DST) strategies [2] in which examples are ordered according to inter-pattern distance, denoted by *delta*. We have investigated three different orders of example presentation using this schedule: *Smallest* difference between patterns (*Smallest Delta Subset Training* (SDST)), where patterns are ordered from the most difficult to the easiest to discriminate; *Largest* difference (*Largest Delta Subset Training* (LDST)), where patterns are ordered from the easiest to the most difficult to discriminate; and *Alternating* difference (*Alternating Delta Subset Training* (ADST)), where the first pattern has the smallest difference, second pattern the largest difference, third pattern the second smallest, etc.

TRAINING STRATEGY HIERARCHY

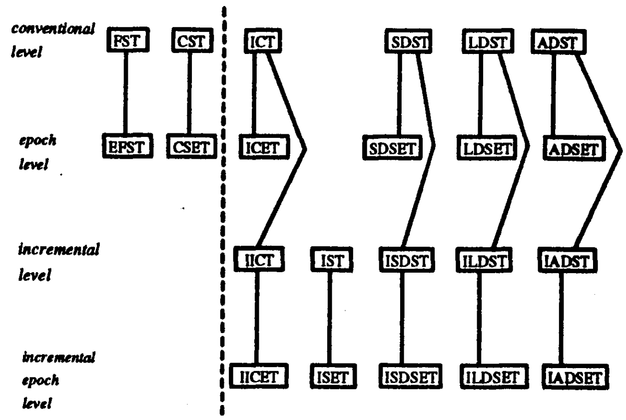


Figure 1. Development of Training Strategies

We further investigate *incremental* construction of subsets, whether complexity ordered or not. In the *incremental* training schedules, e.g. *Incremental Subset Training* (IST) [3], the user specifies an increment (a number of examples) with which each subsequent training subset is enlarged by randomly merging the new training patterns. The initial subset contains the same number of patterns as the increment. Training occurs on the initial subset of training patterns until a success criterion is met, then on the next incrementally constructed subset, etc. For example, when incremental training is combined with ICT, denoted by *Incremental Increased Complexity Training* (IICT) [10], new patterns to be merged are successively taken from the complexity ranked subsets in multiples of the increment, and similarly for DST strategies, which lead to incremental extensions of SDST, LDST, and ADST. For instance, ISDST denotes *Incremental Smallest Delta Subset Training*.

(3) Termination criteria: When training occurs on subsets of the fixed set, training each subset to the same strict precision as required for the fixed set, may cause the network to select a "local" solution instead of correctly generalizing to a global solution. To prevent this from happening we suggest that the initial error value obtained for the untrained network be linearly decreased, or decreased in proportion to the size of the subset and the fixed set, when calculating a required error value for termination of training on a particular subset. In this way the network is gradually restricted to find solutions of higher accuracy.

Furthermore, it is possible that a network can find a good solution without requiring training on the entire fixed set, i.e. a good subset of examples may exist within the fixed set. Therefore we added an outer training loop to the training schedule which terminates further training (and thus subset construction) when the success percentage on a representative test set reaches a user-specified value, thus indicating good generalization and simultaneously preventing overfitting of the data. Testing the network in feedforward mode is usually a quick operation which will not adversely affect the training time, and which has the advantage of detecting when a good set of training patterns has been found by terminating training at that point.

(4) Updating of weights: When using the conventional fixed set training, weights are usually updated after every single pattern, often referred to as "stochastic" weight update, or after presentation of all training patterns in the fixed set, known as updating after an "epoch" (also called "batch updating"). Thus for each of the training strategies already discussed an "epoch update" variation was also investigated in which weights are only updated after presentation of the training subset. For instance, IICET denotes *Incremental Increased Complexity Epoch Training*.

In Figure 1 the hierarchy of training strategies are shown diagrammatically. New strategies are to the right of the vertical dotted line. The conventional level denotes weight updates after every single pattern, the epoch level refers to update after a subset, while the incremental level denotes the incremental construction of subsets.

3 Increased Complexity Training

The idea of *Increased Complexity Training* [9] is simply that it should be quicker to learn the easy part of a problem first, and then gradually increase the complexity of the problem to be learned by giving "more difficult" training patterns in addition to those that the network have already learned. So rather than selecting patterns randomly, patterns are selected "intelligently" by learning easier tasks first. This is reminiscent of human learning – for instance, first learn to count to small numbers, before proceeding to larger numbers.

In this section we demonstrate experimentally that training on increasingly more complex combined subsets reduces drastically the number of weight updates to reach a given Root Mean Square (RMS) error criterion, suggesting that easier subtasks should be learned first and the complexity of the task to be learned gradually increased. Training in this way much reduces the number of attractor basins (possible solutions) thus leading to faster convergence. Convergence time is related to the complexity of the problem, therefore learning easier subproblems first constrains the solution by eliminating many possibilities, causing faster convergence. For ICT and CST, we investigate various schedules of RMS error values to serve as termination criteria for each subsequent subset. We compare these two training strategies with FST for six different ASRNNs using the Counting applications. For ASRNNs we distinguish between single patterns, presented at each time step, and temporal patterns consisting of an entire sequence of single patterns which has to be learned. For all comparisons the initial conditions and parameters were identical for each network, i.e. the same initial weights and the same learning and momentum rates are used.

The Counting application

For the *counting* application [7] an ASRNN sequentially learns to count pulses from zero to 15. The absence of a pulse is represented by a zero, and resets the counter. The network receives one input at a time, and outputs the binary representation of the counter, i.e. four bits. The counting task and Elman ASRNN [5], with one input unit, eight hidden and eight context units, and four output units, are illustrated in Figure 2.

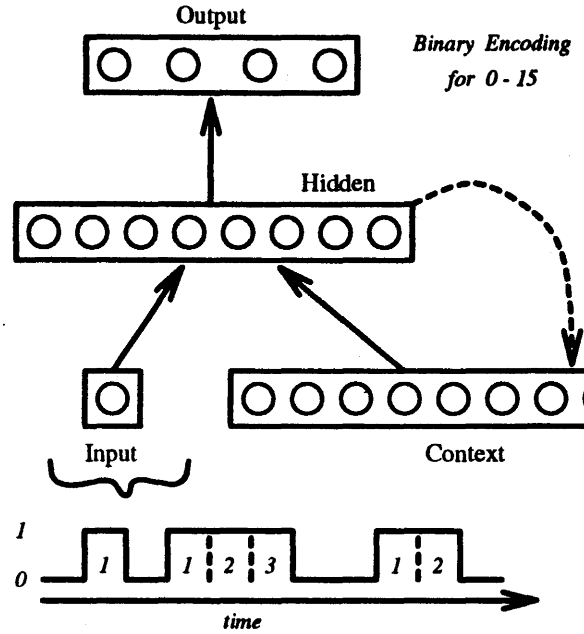


Figure 2. Elman ASRNN for the Counting application

For ICT four training sets, numbered one to four of increasing complexity, were generated as follows: The first set contains two sequential patterns learning to count from 0 to 1, i.e. four single patterns, which requires only the first output unit to vary; the remaining three output a 0. The next training set includes the first in random order and contains sequential patterns to count from 0 to 3, i.e. vary both the first and second output units, and so on for the next two training sets as illustrated in Figure 2. The fixed set contains all 16 temporal patterns (i.e. 137 single patterns) in random order. Four random combined subsets containing 2, 4, 8 and 16 temporal patterns each were generated for experimentation using *CST*.

We have determined the termination criterion for each subsequent subset by investigating various schedules of Root Mean Square error values for ICT and CST. Since the size of the training sets vary, the number of epochs would be an unfair measure for convergence time. The training progress was therefore measured in terms of number of weight updates (i.e. single pattern presentations) required to reach the required RMS termination value. The optimum RMS termination values obtained for ICT and CST respectively are 0.55 – 0.45 – 0.35 – 0.15 and 0.65 – 0.55 – 0.45 – 0.15, where a dash separates the

RMS termination values of each subsequent training subset. These RMS values were obtained out of combinations tried from 0.40 to 0.70 for the first subset, 0.35 to 0.60 for the second, and 0.25 to 0.45 for the third. In every schedule the RMS termination criterion for a particular subset influences the convergence of subsequent training subsets.

For each training strategy, ten simulations were performed with the same initial conditions, except for different initial sets of weights. The average and standard deviation (Std. D.) of the number of updates were determined. The summary of the simulation results appears in Table 1 (the results are compared to FST).

Table 1. A comparison of the average *counting* simulation results for the Elman ASRNN

Training Strategies	Average # updates	Std. D. # updates	Best # updates	% improve (FST)
ICT	17620	6978	9465	44.3%
CST	18260	12116	7878	42.2%
FST	31606	30182	8494	-

From Table 1 it is evident that ICT and CST outperforms FST on average by respectively 44.3% and 42.2%, with ICT being the most consistent by having the lowest standard deviation. The worst case results for FST, CST and ICT were respectively 90968, 53174, and 36487, explaining the large standard deviation for FST.

The second ASRNN for which the training strategies was evaluated, is the Temporal Autoassociation (TA) ASRNN [6], which is designed to reproduce the current input and context units as additional outputs. The same counting experiment was repeated for the TA ASRNN with identical training sets and parameters used for the Elman ASRNN. The TA network consisted of one input unit, eight hidden and eight context units, and 13 output units (4 for representing the binary sum of inputs and the rest for reproducing the inverse mapping). The optimum RMS termination values obtained for ICT and CST were respectively, 0.50 – 0.40 – 0.30 – 0.15 and 0.65 – 0.55 – 0.45 – 0.15. The RMS criterion for each subset for ICT was more strict in every case compared to those for the corresponding CST subsets. This is attributed to the more complex earlier subsets of CST where sets are not increasing in complexity, but random. For each training strategy, ten simulations were performed with the same initial conditions, except for different initial sets of weights. ICT outperformed CST and FST on average by respectively 35.9% (7778 versus 12131 updates) and 46.9% (7778 versus 14645 updates). In spite of the larger network for TA as well as the additional task of learning an inverse mapping of the current input and context units, ICT improved performance by 56% (7778 versus 17620 updates) compared to ICT for the Elman ASRNN. In the next section we compare more thoroughly the performances of the different ASRNNs for counting.

ICT, CST and FST were also evaluated for the Jordan ASRNN [8] with identical training sets and parameters used for the counting task as in the Elman and TA ASRNNs. The Jordan network consisted of one input unit, eight hidden units, four output units; and accordingly four state units (having self-looping connections and one-to-one recurrent connections from the outputs). For Counting ICT and CST outperformed FST on average by respectively 20.7% and 12.5%. The Jordan ASRNN obtained a much larger average number of updates compared to the Elman and TA ASRNN, which is attributed to the former having only four state units as opposed to the eight context units of the latter (implying a less powerful memory) as well as having only an output history at its disposal. ICT showed that although the Jordan ASRNN is not suited to perform well for Counting (this application involves learning input sequences and not output sequences), it still leads to improved performance when compared to the other training strategies.

The training strategies were also evaluated with the Output-to-Hidden Hidden-to-Hidden ASRNN (one of our newly proposed ASRNN architectures), which is designed to incorporate both internal representation (with the Elman network's context units) and output history (with the Jordan network's state units) as input to the network. The ASRNN consisted of one input unit, eight hidden and eight context units, and four output and four state units, with the context and state units operating on the same level as the input. ICT and CST outperformed FST on average by respectively 12.9% and 5.5%. ICT also attained the best number of updates for a particular simulation (4012). The very good simulation results obtained with all the training strategies show that the Output-to-Hidden Hidden-to-Hidden network is indeed a very promising ASRNN. This performance is attributed to the network's ability to deal effectively with input sequences by having a context layer on the input level (similar to the Elman ASRNN).

Another newly proposed ASRNN, the Output-to-Output Hidden-to-Hidden ASRNN, designed to incorporate internal representation history at the input level and output history at the hidden level, was also employed to evaluate the training strategies. The ASRNN consisted of one input unit, eight hidden units, eight context units at the input level, four output units, and four state units operating at the hidden level. From the counting simulation results it were evident that the network found the counting task very difficult to learn and seldom reached the desired RMS value for 0.15. After using a RMS termination value of 0.35 for the final subset, we have obtained optimum RMS termination values of 0.55 – 0.50 – 0.45 and 0.70 – 0.60 – 0.45 for the first three subsets of ICT and CST respectively. Again the RMS criteria for ICT were more strict than those for CST. ICT and CST outperforms FST on average by respectively 26.1% (27034 versus 36579 updates) and 18.5% (29812 versus 36579 updates), where ICT achieved the best number of updates (8077) for a particular simulation. The relatively bad performance of this network compared to the other ASRNNs is attributed to the former having state units operating at the hidden level, which corrupt the internal representation being fed forward to the outputs, since the output history is not useful in remembering the previous temporal input of counting. We conjecture that all ASRNNs with this type of connection would fair poorly

Table 2. A comparison of the different ASRNNs for counting

ASRNNs	Number of network weights	FST # weights updated	FST % improve (Jordan)	CST # weights updated	CST % improve (Jordan)	ICT # weights updated	ICT % improve (Jordan)
O-H H-H	136	873800	86.7%	825384	85.6%	760648	85.4%
TA	176	2577520	60.8%	2135056	62.9%	1368928	73.7%
Elman	104	3287024	50.0%	1899040	67.0%	1832480	64.8%
Jordan	72	6572376	–	5749776	–	5212152	–

for tasks having a temporal input such as counting.

The training strategies were also evaluated with the Output-to-Output ASRNN, which is designed to incorporate only output history at the hidden level. We also test our conjecture that ASRNNs with output-to-output connections would fair badly for tasks having temporal input as opposed to temporal output. The simulation results showed that this ASRNN was unable to learn the counting task, thereby confirming our conjecture that the ASRNN would perform poorly. It is interesting to note that with ICT and CST, the ASRNN initially tried to learn the earlier subsets. For example, for ICT, the RMS value was decreased for the first subset from 0.89 to 0.47, for the second subset from 0.71 to 0.49, for the third from 0.79 to 0.47, and the fourth from 0.82 to 0.71. Although output-to-output connections did not prove to be useful for tasks with temporal input, it should still be useful for tasks having a fixed input and temporal output.

Comparison of different ASRNNs for Counting

In this section we compare the performance of different ASRNNs for the counting task. For a fair comparison we have to multiply the average number of updates (single pattern presentations) by the number of weights in each network to obtain the total number of weights updated. Since the desired RMS termination value for the final subset of the Output-to-Output Hidden-to-Hidden ASRNN was not the same as the one used for the other ASRNNs and the simulations for the Output-to-Output ASRNN did not converge, we only compare the Elman, Temporal Autoassociation, Jordan, and Output-to-Hidden Hidden-to-Hidden ASRNNs for each training strategy. The comparison results appear in Table 2.

It follows from Table 2 that the Output-to-Hidden Hidden-to-Hidden (O-H H-H) ASRNN outperforms the Jordan, Elman, and TA ASRNNs for FST by respectively 86.7%, 73.4% , and 66.1%. The respective percentages for O-H H-H's ICT are 85.4%, 58.5%, and 44.4%, and for CST are 85.6%, 56.5% and 61.3%. Compared to the Jordan ASRNN, the O-H H-H ASRNN improved the number of weights updated for all three training strategies consistently by more than 85%. Although the TA ASRNN improved the performance of FST and ICT when compared to the Elman ASRNN, the opposite is true for CST. From this counting results we assert that the newly proposed Output-to-Hidden Hidden-to-Hidden ASRNN indeed distinguished itself as the best ASRNN. Since the O-H H-H ASRNN incorporates the history of the internal representations as well as the output, causing this network to be able to learn temporal input and/or output sequences, we conjecture that this general-purpose ASRNN would be able to learn most temporal tasks of this kind very efficiently.

4 Incremental Training Strategies

In this section we elaborate on the two new incremental training strategies, *Incremental Subset Training* and *Incremental Increased Complexity Training*. IST also caters for the fact that not all training sets can be partitioned into increased complexity subsets. These strategies are compared with ICT, CST, and FST for ASRNNs and Feedforward BP networks using the Addition and Cluster detection applications. A theoretical lower bound is given for the least number of updates when following an incremental training schedule. For the correct increment in subset size the number of updates required for Addition almost attained the predicted lower bound.

The training schedules of ICT and CST leave room for much variation in the size of the initial subset, the increment in subset size, termination criteria for each subsequent subset, and whether training can be terminated earlier without learning on the entire fixed set. This last point is desirable since good generalization may be obtained on a subset of the fixed set. We therefore investigated the following incremental training schedule: As success criterion on a subset of examples the RMS error value is used to take into account that subset sizes increase. The user specifies the desired final RMS error value (*RMSdesired*) which signifies successful training, as well as the increment in subset size (*SUBincrement*), i.e. the number of single or temporal patterns (for respectively feedforward networks and ASRNNs) to be randomly added to the training subsets. If an independent test set is available, the user specifies the percentage of test examples which should be correctly given by the fully trained network in feedforward mode. This allows that training be terminated as soon as the network has generalized sufficiently. The initial subset size is the same as the subset increment. The training schedule then consists of two nested loops: The outer loop repeats until successful generalization is obtained on the test set, while the inner loop repeats training on the subset until the required RMS error value for that training subset is obtained. The RMS termination values for each training subset is decremented linearly (by *RMSdecrement*) from that obtained for the initial subset (*RMSstart*) to the final user-specified value (*RMSdesired*) divided by the number of RMS decrements, thus gradually requiring stricter termination criteria. This prevents training to a too high accuracy on a small set of examples, which may cause an inferior solution to be selected instead of generalizing sufficiently. When the desired RMS value for a particular subset has been reached, the next subset is constructed, the new RMS value is computed and the process repeated. The number of incremental subsets (and number of RMS decrements) is

$$S = \left\lceil \frac{\text{Number of training patterns in fixed set } (\textit{PATtotal})}{\text{Subset increment } (\textit{SUBincrement})} \right\rceil$$

For ASRNNs *PATtotal* is the number of temporal training patterns in the fixed set, whereas for feedforward networks it is the number of single patterns. *RMScurrent* denotes the current RMS error value obtained by the network after presenting the current subset (*SUBcurrent*). The desired RMS value for the current subset is *RMSstart* - ((*SUBcurrent* - 1) * *RMSdecrement*). IST constructs subsequent incremental training subsets by randomly adding training patterns, whereas IICT constructs the subsets by adding patterns successively from the complexity ranked subsets in multiples of the increment.

To obtain a theoretical lower bound for the number of weight updates with feedforward networks and ASRNNs, let us denote the subset increment for single and temporal patterns respectively by I_s and I_t. Let us further assume that weights are updated after every single pattern. If training is not repeated on any subset when single patterns are used, then the lower bound for the number of weight updates on the entire training set is $I_s S(S+1)/2$ since the incremental subset sizes are $I_s, 2I_s, \ldots, SI_s$. However, when the training set consists of temporal patterns and the same assumptions hold, the lower bound for the number of updates on the entire training set is $LI_t S(S+1)/2$, where L denotes the average length of the temporal patterns and the incremental subset sizes are $I_t, 2I_t, \ldots, SI_t$. In the following section we show how the lower bound formula for temporal patterns could be used to predict the lowest number of updates when only a portion of the incremental subsets are needed to reach the desired termination criteria.

The Addition application

For the *addition* experiment [4] the aim is to learn to sequentially add two base four numbers. Each base four number is given a two-digit binary representation. The Elman ASRNN, illustrated in Figure 3, is confined to one column of digits at a time. It has five inputs, four representing the one column of digits and one indicating the end of the input, 16 hidden and 16 context units, and six output units representing the sum (two units) of the one column of digits and the four possible actions (four units). Actions are to write the sum, to remember or output the carry, shift to the next column of digits, and indicate if done. The addition task is very complex, because not only must it learn a sequence of inputs, but for each input a different sequence of outputs is required.

For one-, two- and three-column addition there are 4096 different temporal patterns. The training set for FST

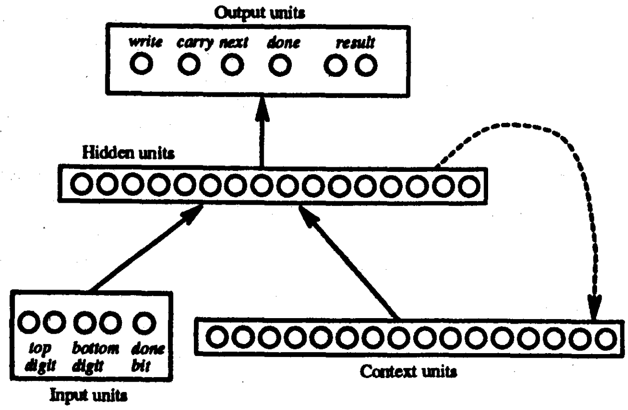


Figure 3. Elman ASRNN for the Addition application

consisted of 8% of this set, constituting 320 temporal patterns, whereas the independent test set consisted of the other 92% of the total set. For ICT three training subsets of increasing complexity were generated, corresponding to 1%, 4% and 8% of the total training set, and each set containing patterns for adding, respectively, one column of figures, two and three columns. For CST the three subsets contained the same number of temporal patterns as each ICT subset, but randomly chosen. Again all network initial conditions for each experiment were identical. For comparison purposes, the final subsets of ICT and CST were the same one used for FST.

The following two termination criteria were used: A RMS error value ≤ 0.15 for the final subset, and a success percentage on independent test set of $\geq 97\%$. The optimum RMS termination values obtained for ICT and CST were both $0.65 - 0.45 - 0.15$. For the incremental training strategies, IST and IICT, experiments were repeated with increments of 1, 2, 4, 8, 16 and 32 temporal patterns with corresponding linear RMS decrements. The results of ten simulations for each training strategy with the same initial conditions, except for different initial weights, are summarized in Table 3.

Table 3. A comparison of the average *Addition* simulation results for the Elman ASRNN

Training Strategies	Average # updates	Std. D. # updates	Best # updates	% improve (FST)
IICT	32703	11951	16292	28.7%
IST	34700	11220	22582	24.4%
ICT	36186	9388	22735	21.1%
CST	39478	20579	25726	13.9%
FST	45870	16312	29965	-

From Table 3 it is apparent that IICT outperforms FST, CST, and ICT on average by respectively 28.7%, 17.2% and 9.6%, with IICT accordingly obtaining the best number of updates (16292). IST improved the average number of updates compared to FST, CST, and ICT by respectively 24.4%, 12.1%, and 4.1%.

Since the lower bound for the number of updates on the total number of subsets S is $LI_t S(S+1)/2$, the corresponding lower bound for the number of updates on a portion of the subsets, say S_p, is $LI_t S_p(S_p+1)/2$. For increments 1, 2, 4 and 8 the number of updates required almost attained the optimum lower bound. The number of updates required for IST (with increment four) was 22582, whereas the calculated lower bound was 22464 for an average temporal pattern length of 7.2, and requiring only 156 of the 320 temporal patterns (39 subsets out of a possible 80). When this set of 156 patterns were used for FST, 45360 updates were necessary compared to the 22582 of IST to also reach 97% success on the test set. For increments 16 and 32 more updates were required due to the repetition of their final subsets to reach the desired RMS criteria and generalization. Table 4 contains the detailed results for IST and FST.

Table 4. Training results for IST and FST

IST: I_t	S	Updates	S_p	Patterns	Lower Bound
1	320	141631	197	197	140422
2	160	23075	56	112	22982
4	80	22582	39	156	22464
8	40	36620	35	280	36288
16	20	28975	20	320	24192
32	10	61420	10	320	12672
FST	1	48405	1	320	—

The erratic behaviour of the error curve for BP learning prompted the investigation leading to the conclusion that BP is sensitive to initial conditions. The proposed incremental training strategies reduce the effect of many attractor basins which cause the erratic behaviour of the error curve, thus leading to faster convergence.

The Addition experiment with the incremental training strategies was repeated for the Temporal Autoassociation ASRNN. The simulation results showed that the addition experiment never converged for the TA ASRNN, no matter what training strategy was used. It was evident that the two incremental training strategies performed the most respectable under the circumstances. The TA ASRNN is not a suitable architecture for the addition problem, since the latter requires a many to one mapping to be learned, whereas the TA network demands unique internal representations for the different inputs.

IST and IICT were also evaluated for the Jordan ASRNN. The optimum subset increment obtained for IICT and IST were respectively four and six. IICT performing remarkably well on average (52044 updates) when compared to FST, CST, IST, and ICT. It improved the average number of weight updates by respectively 61.8%, 61.5%, 59.5% and 57.6%. The importance of training with subsets that increase in complexity is not only emphasized by IICT's performance, but also by the fact that even ICT outperformed IST by 4.5%. When comparing the Elman and Jordan ASRNN's performance for IST and IICT (with the network weights taken into account), we obtain respective improvements for the Elman ASRNN of 57.1% and only 0.2%, further accentuating IICT's excellent performance for the Jordan ASRNN.

The Cluster Detection application

The training strategies are evaluated here for a feedforward BP network using the Cluster detection experiment, where the aim is to detect two classes in four clusters of input, as illustrated in Figure 4. The feedforward BP network

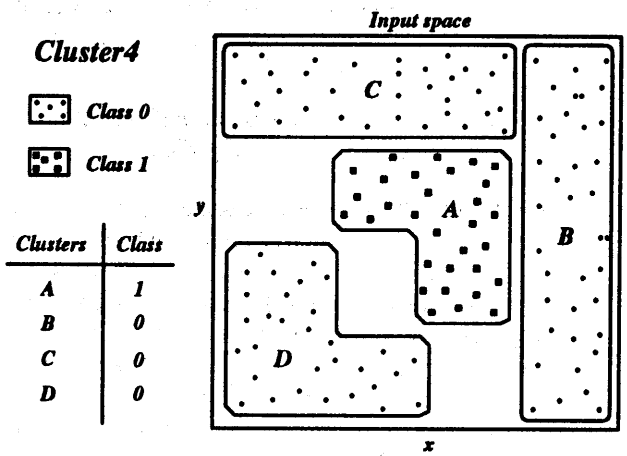


Figure 4. The Cluster Detection experiment: Detecting two classes in four clusters of input

consisted of two inputs, representing the x and y coordinates of a point in the input space, three hidden units, and two output units, representing a class of type zero or one. A total set of 1300 points were generated from the four clusters in the input space. The training set for FST consisted of 20% of this set, constituting 260 input patterns (130 patterns from each class), whereas the independent test set consisted of the other 80% of the total set (1040 single patterns). For ICT three training subsets of increasing complexity were generated, corresponding to clusters $A\&B$ (94 single patterns), clusters $A\&B\&C$ (176 single patterns) and clusters $A\&B\&C\&D$ (260 single patterns). For CST the three subsets contained the same number of single patterns as each ICT subset, but randomly chosen from the four clusters. Epoch updating was also evaluated, and the optimum subset increment obtained for IST, ISET, IICT, IICET were respectively 2, 2, 1, and 6. From Table 5

Table 5. A comparison of the average Cluster detection simulation results for the Feedforward BP network

Training Strategies	Average # updates	Std. D. # updates	Best # updates	% improve (FST)
IICT	30864	3108	26996	45.4%
IST	31652	4073	29446	44.0%
ICT	30199	3945	25314	46.6%
CST	34395	4206	26734	39.2%
FST	56550	19339	31200	-

it appears that the average performance (of 10 simulations) of the incremental training strategies are on par with that of ICT, and also outperformed FST and CST: IICT improved performance by respectively 45.4% and 10.3%, and IST by 44% and 8%. IICT's average number of updates is the most consistent by having the lowest standard deviation (3108).

From Table 6 it is evident that better results are obtained with epoch updating when the patterns are ordered

Table 6. A comparison of the average Cluster detection simulation results for the Feedforward BP network – epoch updating

Training Strategies	Average # updates	Std. D. # updates	Best # updates	% improve (FST)
IICET	26286	6541	21365	53.5%
ISET	32071	17005	20020	43.3%
ICET	22035	10783	14526	61.0%
CSET	36909	17606	20296	34.7%
FSET	38740	7920	29900	31.5%

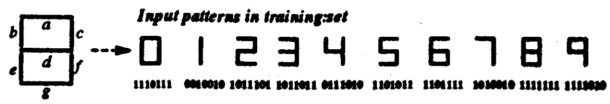
in an increased complexity fashion as opposed to randomly ordered. For example, ICET and IICET, improved performance compared to ICT and IICT by respectively 27% and 14.8%, whereas CST and IST outperformed CSET and ISET by respectively 6.8% and 1.3%.

5 Delta Training Strategies

In this section we describe and evaluate a group of *Delta Training Strategies*, which is developed independently of the problem's complexity, since the order of its training patterns is based on inter-pattern distance, denoted by *delta*. Any suitable distance metric, such as Hamming or Euclidean distance, can be used. A pattern, in this context, can be a single or temporal input pattern. We employ a method, called the *Delta Ranking Method* (DRM), which determine the complexity relation between the input patterns by obtaining their inter-pattern distances and then ranking them according to some scheme. We explain the method by using the example illustrated in Figure 5 where each one of the digits 0...9 is displayed as a seven-bit single input pattern. The first step is to determine the minimum distances between all the input patterns in the set. For example, the minimum distance between the digits 2 and 4 is five, since their input patterns differ by five bits. We then obtain a square symmetric matrix (with diagonal coefficients of zero) which consists of delta values as illustrated in the figure. The second step is to compute the *DRM score*, which is the sum of all the minimum distances for each input pattern. This is achieved by adding up the delta values in each row or column. The third step is then to rank the input patterns according to some ranking scheme, such as from easiest to most difficult to discriminate. For example, since we have obtained the largest sum of minimum distances (36) for digit 1, it implies that this input pattern is the easiest to discern.

We have investigated three different orders of presenting the training patterns using the Delta Ranking Method. The first training strategy, called *Smallest Delta Subset Training* (SDST), orders the input patterns according the smallest DRM score between input patterns, i.e. patterns are ordered from the most difficult to the easiest to discriminate. The next strategy, *Largest Delta Subset Training* (LDST), involves an order presentation of patterns with largest DRM score presented first, i.e. patterns are ordered from the easiest to the most difficult to discriminate. The final strategy is called the *Alternating Delta Subset Training* (ADST), where the first pattern has the smallest DRM score, the second pattern the largest DRM score, third pattern the second smallest, etc.

DELTA RANKING METHOD (DRM) - an example



STEP 1:

Determine the minimum inter-pattern distance (delta) between all the input patterns in the set.

	0	1	2	3	4	5	6	7	8	9
0	0	4	3	3	3	3	2	3	1	3
1	4	0	6	3	2	6	6	1	5	3
2	3	6	0	2	5	2	3	4	2	4
3	3	3	2	0	4	4	3	2	2	3
4	3	2	5	4	0	4	4	3	3	1
5	3	6	2	4	4	0	1	4	2	2
6	2	6	3	3	4	1	0	5	1	3
7	3	1	4	2	3	4	5	0	4	2
8	1	5	2	2	3	2	1	4	0	2
9	3	3	4	3	1	2	3	2	2	0

STEP 2:

Compute the DRM score, i.e., the sum of all the minimum distances for each input pattern.

Pattern	DRM
0	25
1	36
2	31
3	26
4	29
5	28
6	28
7	28
8	22
9	23

STEP 3:

Rank the input patterns according to some ranking scheme, say from easiest to most difficult to discriminate.

Input pattern order
1
2
4
5 6 7
3
0
9
8

Figure 5. Illustration of Delta Ranking Method for Delta Training Stategies

We have further investigated the incremental extensions of SDST, LDST, and ADST, which employ the same training schedule as IST and IICT, and respectively led to *Incremental SDST* (ISDST), *Incremental LDST* (ILDST), and *Incremental ADST* (IADST). These incremental delta training strategies construct the subsets by adding patterns successively from the DRM score ranked subsets in multiples of the increment and according to one of the proposed ranking schemes. For feedforward networks we have also evaluated the epoch updating versions of all six these strategies.

The Delta Ranking Method can also be applied to temporal patterns, where a DRM score is then computed for each different temporal pattern in the training set. The DRM can in general be applied to any set of n-dimensional real valued input patterns, since the basic computation only involves determining the minimum distance between a point $X = (x_1, x_2, \ldots, x_n)$ and $Y = (y_1, y_2, \ldots, y_n)$. Assume we have N training input patterns. The computational complexity for the first step of DRM is then $N(N-1)/2$, since there are $1 + 2 + \ldots + (N-1)$ computations in determining the minimum distances between the N input patterns. The complexity of computing the second step is $N(N-1)$, since there are N input patterns, for which a DRM score, consisting of $N-1$ additions, must be computed.

The Digital Morse Code Generation application

We first evaluated the delta training strategies for the Elman ASRNN using the Digital Morse code generation application, where the task is to sequentially generate the Morse code sequence for each of the 10 decimal digits, presented as a seven-bit input pattern. The network, illustrated in Figure 6, has seven input units, nine hidden and nine context units (optimum number experimentally determined),

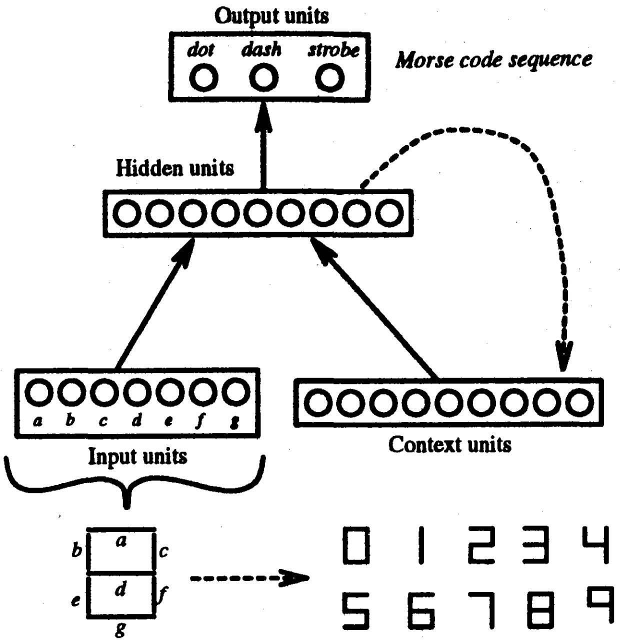


Figure 6. Elman ASRNN used for the Digital Morse code generation application

and three binary outputs respectively indicating *dot*, *dash* and a *done* bit which is on only when the last dot or dash is output.

For successful training a RMS error value of less that 0.15 for the training set and a success ratio greater than 98% on the test set were required. For SDST, LDST, and ADST four subsets were constructed according to the ranking schemes described in the previous section. The optimum RMS termination values for these three training strategies were respectively, 0.75 – 0.65 – 0.55 – 0.15, 0.55 – 0.50 – 0.40 – 0.15, and 0.65 – 0.55 – 0.45 – 0.15. These values make sense, since they were less strict to accomodate the more complex subsets of SDST, very strict for the easier subsets of LDST, and in between for the alternating patterns of ADST. The optimum RMS termination values for CST were 0.65 – 0.55 – 0.45 – 0.15. The optimum subset increment obtained for ISDST, ILDST, IADST, and IST were respectively two, three, five, and one. The average simulation results of ten simulations for each training strategy are summarized in Table 7.

Table 7. A comparison of the average *Digital Morse Code Generation* simulation results for the Elman ASRNN

Training Strategies	Average # updates	Std. D. # updates	Best # updates	% improve (FST)
ILDST	30870	9731	15465	58.1%
ISDST	32646	20769	12690	55.7%
IADST	71682	24252	28450	2.7%
IST	45346	25922	16950	38.5%
SDST	65580	31182	24600	11.0%
LDST	52960	29399	14500	28.1%
ADST	89380	4860	74800	-
CST	69730	26692	29650	5.4%
FST	73700	27398	14500	-

From Table 7 ILDST and ISDST outperformed all the other training strategies on average, including IST. ILDST compared to FST, CST, SDST, LDST, IST, and ISDST yielded improvements of 58.1%, 55.7%, 53.1%, 41.7%, 32.1%, and 5.4%. The optimum increment in patterns for ILDST, ISDST, and IST were small (respectively 3, 2, and 1), suggesting in general that small increments perform better. This is supported by the observation that single pattern updating rather than epoch updating usually causes faster convergence for ASRNNs. We conjecture that the Incremental Delta Training strategies (which start with a small set of examples) allow weights to find a good position in weight space quickly, thus requiring less weight movement than that suggested when the entire training set is presented at once.

Digit Recognition

The delta training strategies were finally tested on a Digit recognition application for a feedforward BP network. The task comprises the recognition of binary patterns representing the 10 decimal digits encoded for a seven-segment display. The input patterns were similar to the input of the Digital Morse code generation task, except that at most one bit of a pattern was randomly corrupted with 10% probability. The network has seven input units, ten hidden units (the optimum value experimentally determined), and one real valued output. For successful training a RMS error value of less than 0.185 for the training set and a success percentage greater than 80% on the test set were required. The fixed training set had 349 patterns and the test set 151.

The optimum RMS termination values were again obtained for the four subsets of SDST (SDSET), LDST (LDSET), ADST (ADSET), and CST (CSET). The optimum subset increment obtained for ISDST, ISDSET, ILDST, ILDSET, IADST, IADSET, IST, and ISET were respectively 2, 5, 1, 4, 11, 12, 3, and 5. We note that the optimum subset increment is always larger for the epoch version of a particular training strategy, suggesting that for epoch updating to be effective the subset must rather be larger than smaller. The average simulation results of ten simulations for each non-epoch updating training strategy are summarized in Table 8.

Table 8. A comparison of the average *Digit Recognition* simulation results for the Elman ASRNN

Training Strategies	Average # updates	Std. D. # updates	Best # updates	% improve (FST)
ISDST	21274	5968	14524	64.5%
IADST	29149	10457	19453	51.3%
ILDST	32951	4615	24531	45.0%
IST	27463	6290	20343	54.1%
ADST	30260	5522	19023	49.5%
LDST	33848	4197	27217	43.5%
SDST	36562	7438	26267	38.9%
CST	53607	11135	30254	10.5%
FST	59888	15274	33504	-

For this application ISDST performed the best among the non-epoch training strategies, improving the average updates compared to FST, CST, LDST, SDST, ADST, IST, IADST, and ILDST by respectively 64.5%, 51.9%, 48.6%, 41%, 33.9%, 19.8%, 12.8%, and 8.1%. The very good performance of the ISDST can be attributed to the good discriminating position in weight space achieved by presenting the network initially with a few complex patterns in an incremental fashion. We note that ADST as a non-incremental delta strategy, obtained the second best number of updates. For this application all the non-epoch delta training strategies outperformed the conventional FST and CST strategies.

Table 9. A comparison of the average *Digit Recognition* simulation results for the Elman ASRNN - epoch updating

Training Strategies	Average # updates	Std. D. # updates	Best # updates	% improve (FST)
ISDSET	19926	5688	14111	66.7%
ILDSET	27161	4585	21943	54.6%
IADSET	34191	10983	15675	42.9%
ISET	24694	6110	16630	58.8%
ADSET	29664	3933	21742	50.5%
LDSET	31998	3274	26519	46.6%
SDSET	31301	3605	27427	47.7%
CSET	42381	16148	25213	29.2%
FSET	49139	20807	24081	17.9%

From Table 9 it is evident that better results are obtained with epoch updating (the results are compared to FST). ISDSET improved performance compared to FST, FSET, CSET, SDSET, LDSET, ADSET, ISET, IADSET, and ILDSET by respectively 66.7%, 59.4%, 53%, 36.3%, 37.7%, 32.8%, 19.3%, 41.7%, and 26.6%. It also obtained the best number of updates, namely 14111. Again all the Delta training strategies proved to be very effective in reducing the training time, especially when compared with the conventional strategies CSET and FSET. Thus the ordering of training patterns according to the proposed schemes, especially in an incremental fashion, allows that a good discriminating position in weight space be the target early in the learning process, yielding faster convergence.

6 Conclusion

In this paper we have showed that the variation of a training strategy is a viable alternative method to produce an effective solution for training ASRNNs and feedforward networks within a feasible time. We have developed *Increased Complexity Training* (ICT), two *incremental* training strategies (IICT and IST) and six *delta* training strategies (SDST, LDST, ADST, ISDST, ILDST, and IADST), and showed that they lead to considerable improvement in training time and alleviate problems such as a too large training set and a long training time. Training strategies and ASRNNs were evaluated simultaneously for different applications, varying in complexity and ranging from sequence recognition to sequence generation.

It was demonstrated with six different ASRNNs and feedforward networks that for several applications, ICT outperformed the conventional CST and FST. The incremental training strategies also effectively addressed further problems with the size of the initial subset, the increment in the subset size, the termination criteria for each subsequent subset, and the number of examples required for good generalization. IST and IICT showed for Addition that a good training set can be quite small to provide very good generalization. IST, for example, for a particular simulation needed only half of the training set to double the performance of training on a fixed set. It also improved performance by 50% compared to FST on the same half of the training set. For the correct increment in subset size the number of updates required for addition almost attained the theoretical lower bound.

The delta training strategies employ the Delta Ranking Method, which determines the complexity relation between the input patterns by obtaining their inter-pattern distances and then ranking them according to some scheme. Three basic ranking schemes were introduced that led to the different delta training strategies. Incremental delta training strategies performed the best overall, signalling that the ordering of training patterns according to our proposed delta ranking schemes, especially when presented in an incremental fashion, forces the network to discriminate between classes early in the training process, leading to reduced training time.

We have also evaluated Elman, Jordan, TA and other newly constructed ASRNNs for the Counting and Addition tasks, and found the Output-to-Hidden Hidden-to-Hidden ASRNN to be the best simple recurrent network architecture, outperforming the other ASRNNs between 44% and 87% for the training strategies. This network combines the best features of the Elman and Jordan ASRNN, causing it to be able to learn input and/or output sequences.

References

1. I Cloete and J Ludik. 'Increased complexity training'. In J C J Mira and A Prieto, eds., *New Trends in Neural Computation*, pp. 267–271. Springer-Verlag, Berlin, (1993).
2. I Cloete and J Ludik. 'Delta training strategies'. In *Proceedings of the IEEE World Congress on Computational Intelligence (WCCI'94)*, pp. 1:295–298, Orlando, Florida, USA, (26 June to 2 July 1994).
3. I Cloete and J Ludik. 'Incremental training strategies'. In *Proceedings of the International Conference on Artificial Neural Networks (ICANN'94)*, pp. 2:743–746, Sorrento, Italy, (26-29 May 1994).
4. G W Cottrell and F S Tsung. 'Learning simple arithmetic procedures'. In J A Barnden and J B Pollack, eds., *High-Level Connectionist Models*, pp. 1:305–321. (1991).
5. J L Elman. 'Finding structure in time'. *Cognitive Science*, **14**:179–211, (1990).
6. Z Ghahramani and R B Allen. 'Temporal processing with connectionist networks'. In *Proceedings of the International Joint Conference on Neural Net-*

works (IJCNN'91), pp. 2:541–546, Seattle, Washington, USA, (1991).

7. J Hoekstra. 'Is counting with artificial neural networks a problem?'. In *Proceedings of the 1st IFIP Working Group-10.6 Workshop*, pp. 27–30, Grenoble, France, (2-3 March 1992).
8. M I Jordan. 'Attractor dynamics and parallelism in a connectionist sequential machine'. In *Proceedings of the 8th Conference on Cognitive Science*, pp. 531–546, Amherst, MA, USA, (1986).
9. J Ludik and I Cloete. 'Training schedules for improved convergence'. In *Proceedings of the International Joint Conference of Neural Networks (IJCNN'93)*, pp. 1:561–564, Nagoya, Japan, (October 1993).
10. J Ludik and I Cloete. 'Incremental increased complexity training'. In *Proceedings of the European Symposium on Artificial Neural Networks (ESANN'94)*, pp. 161–165, Brussels, Belgium, (20-21 April 1994).
11. F M Silva and L B Almeida. 'Speeding up backpropagation'. In R Eckmiller, ed., *Advanced Neural Computers*. Elsevier Science Publishers, Amsterdam, (1990).

Appendix A Nomenclature

ASRNN	Architecture-specific Recurrent Neural Network
ADST	Alternating Delta Subset Training
ADSET	Alternating Delta Subset Epoch Training
CST	Combined Subset Training
DRM	Delta Ranking Method
DST	Delta Subset Training
FST	Fixed Set Training
FSET	Fixed Set Epoch Training
ICT	Increased Complexity Training
ICET	Increased Complexity Epoch Training
IICT	Incremental Increased Complexity Training
IICET	Incremental Increased Complexity Epoch Training
IST	Incremental Subset Training
ISET	Incremental Subset Epoch Training
IADST	Incremental Alternating Delta Subset Training
ILDST	Incremental Largest Delta Subset Training
ISDST	Incremental Smallest Delta Subset Training
IADSET	Incremental Alternating Delta Subset Epoch Training
ILDSET	Incremental Largest Delta Subset Epoch Training
ISDSET	Incremental Smallest Delta Subset Epoch Training
LDST	Largest Delta Subset Training
LDSET	Largest Delta Subset Epoch Training
SDST	Smallest Delta Subset Training
SDSET	Smallest Delta Subset Epoch Training
TA	Temporal Autoassociation

Received: 2/95, Accepted: 4/95, Final copy: 6/95.

Notes for Contributors

The prime purpose of the journal is to publish original research papers in the fields of Computer Science and Information Systems, as well as shorter technical research notes. However, non-refereed review and exploratory articles of interest to the journal's readers will be considered for publication under sections marked as Communications or Viewpoints. While English is the preferred language of the journal, papers in Afrikaans will also be accepted. Typed manuscripts for review should be submitted in triplicate to the editor.

Form of Manuscript

Manuscripts for *review* should be prepared according to the following guidelines.

- Use wide margins and $1\frac{1}{2}$ or double spacing.
- The first page should include:
  - title (as brief as possible);
  - author's initials and surname;
  - author's affiliation and address;
  - an abstract of less than 200 words;
  - an appropriate keyword list;
  - a list of relevant Computing Review Categories.
- Tables and figures should be numbered and titled.
- References should be listed at the end of the text in alphabetic order of the (first) author's surname, and should be cited in the text in square brackets [1–3]. References should take the form shown at the end of these notes.

Manuscripts accepted for publication should comply with the above guidelines (except for the spacing requirements), and may be provided in one of the following formats (listed in order of preference):

1. As (a) LaTeX file(s), either on a diskette, or via e-mail/ftp – a LaTeX style file is available from the production editor;
2. As an ASCII file accompanied by a hard-copy showing formatting intentions:
   - Tables and figures should be original line drawings/printouts, (not photocopies) on separate sheets of paper, clearly numbered on the back and ready for cutting and pasting. Figure titles should appear in the text where the figures are to be placed.
   - Mathematical and other symbols may be either handwritten or typed. Greek letters and unusual symbols should be identified in the margin, if they are not clear in the text.

   Contact the production editor for markup instructions.
3. In exceptional cases camera-ready format may be accepted – a detailed page specification is available from the production editor;

Authors of accepted papers will be required to sign a copyright transfer form.

Charges

Charges per final page will be levied on papers accepted for publication. They will be scaled to reflect typesetting, reproduction and other costs. Currently, the minimum rate is R30-00 per final page for LaTeX or camera-ready contributions that require no further attention. The maximum is R120-00 per page (charges include VAT).

These charges may be waived upon request of the author and at the discretion of the editor.

Proofs

Proofs of accepted papers in category 2 above may be sent to the author to ensure that typesetting is correct, and not for addition of new material or major amendments to the text. Corrected proofs should be returned to the production editor within three days.

Note that, in the case of camera-ready submissions, it is the author's responsibility to ensure that such submissions are error-free. Camera-ready submissions will only be accepted if they are in strict accordance with the detailed guidelines.

Letters and Communications

Letters to the editor are welcomed. They should be signed, and should be limited to less than about 500 words.

Announcements and communications of interest to the readership will be considered for publication in a separate section of the journal. Communications may also reflect minor research contributions. However, such communications will not be refereed and will not be deemed as fully-fledged publications for state subsidy purposes.

Book reviews

Contributions in this regard will be welcomed. Views and opinions expressed in such reviews should, however, be regarded as those of the reviewer alone.

Advertisement

Placement of advertisements at R1000-00 per full page per issue and R500-00 per half page per issue will be considered. These charges exclude specialized production costs which will be borne by the advertiser. Enquiries should be directed to the editor.

References

1. E Ashcroft and Z Manna. 'The translation of 'goto' programs to 'while' programs'. In *Proceedings of IFIP Congress 71*, pp. 250–255, Amsterdam, (1972). North-Holland.
2. C Bohm and G Jacopini. 'Flow diagrams, turing machines and languages with only two formation rules'. *Communications of the ACM*, **9**:366–371, (1966).
3. S Ginsburg. *Mathematical theory of context free languages*. McGraw Hill, New York, 1966.

South African Computer Journal

Number 14, July 1995
ISSN 1015-7999

Suid-Afrikaanse Rekenaar-tydskrif

Nommer 14, Julie 1995
ISSN 1015-7999

Contents