Learning through experience: an emergent connectionist account of letter production behaviour

Fiona M. Richardson [1,2], Neil Davey[1], John Done[2] & Rod Adams[1]

F.1.Richardson, N.Davey, D.J.Done, R.G.Adams@herts.ac.uk

Department of Computer Science[1], Department of Psychology[2],

University of Hertfordshire, College Lane, Hatfield, Hertfordshire, AL10 9AB, UK

Word count: 5986

Address for correspondence:

Dr. Fiona Richardson
School of Psychology
Birkbeck College, University of London
Malet Street, Bloomsbury
London WC1E 7HX, UK
Email: f.richardson@bbk.ac.uk
Tel.:   +44 (0)20 7631 6468
Fax: +44 (0)20 7631 6312

Abstract

In handwriting the drawing or copying of an individual letter involves a process of

<u>linearising</u> whereby the form of the letter is broken down into a temporal sequence of

strokes for production. In experienced writers, letters are produced consistently using

the same production methods that are economic in terms of movement. This regularity

permits a rule-based description of such production processes, which can be used in

the teaching of handwriting skills. In this paper, the outstanding question from rule-

based descriptions as to how consistent and stable letter production behaviour

emerges as a product of practice and experience is addressed through the

implementation of a connectionist model of sequential letter production. This model:

(1) examines the emergence of letter production behaviour, namely - the linearising

process, (2) explores how letters may be internally represented across both spatial and

temporal dimensions, and (3) investigates the impact of learning certain letter

production methods when generalising to produce novel letterforms. In conclusion,

the connectionist model offers an emergent account of letter production behaviour,

which addresses the co-representation of spatial and temporal dimensions of letters,

and the impact of learning experiences upon behaviour.

1.   Introduction

Handwriting involves the structured sequencing of movements in order to produce a letter or a string of letters. The process whereby a complete static letterform is broken down into a temporal sequence of strokes for production is known as *linearising* (Thomassen and Tibosh 1991, Thomassen, Meulenbroek, and Tibosh 1991). This fundamental initial stage in the handwriting process is relatively autonomous, and is likely to be subject to a variety of influences such as culture and handedness, in combination with the level of skill and practice (Alston and Taylor 1987, 1988, Thomassen and Tibosh 1991, Thomassen, Meulenbroek, and Tibosh 1991, Meulenbroek and Thomassen 1993, Desbiez, Vinter, and Meulenbroek 1996a, b). Generally, the sequence in which individual letter strokes are produced is strongly influenced by the direction in which the writing will be read (Alston and Taylor 1990). For writing that is produced from left to right (as in Western cultures), the sequence of strokes usually commences at the leftmost point of the letter, progressing through neighbouring strokes, although in some cases letters are produced starting with the topmost point of the first vertical line of the letter.

In the experienced writer, letter production is fluent and economical (Thomassen, Meulenbroek, and Tibosh 1991). Letters are produced in a manner that is both biomechanically and cognitively efficient, and is typified by a production trajectory that uses the minimum number of strokes, with ideally the movement trajectory being as continuous (as few pen-up movements and positional changes) as possible (Meulenbroek and Thomassen 1991, 1993, Meulenbroek, Thomassen, Schillings and Rosenbaum 1996, Desbiez, Vinter, and Meulenbroek 1996a, 1996b). This smooth production movement, (for which visual feedback during production is not a requirement: Marquardt, van Doorn and Keuss 1993, Gentz and Mai 1996)

forms a characteristic bell-shaped velocity profile. Such learnt letter production trajectories are considered to be stored as higher-level representations, which are essentially motor programs for movement (Teulings and Shoemaker 1993, van Galen and Webber 1998). In this respect, economical production processes are also consistent production processes, with the same letters being produced repeatedly in a similar manner (Thomassen and Tibosh 1991). Furthermore, as similar letters will not have radically different biomechanical and cognitive costs, consistency can also be observed when comparing similar individual letters. For example, the curved segment/s of letters P, R, B, and D, are all typically produced in a clockwise direction. These regularities of letter production processes are reflected in educational models used in the teaching of handwriting (for examples, see Alston and Taylor 1990, Armstrong 1993, Hadley 1996), which are usually very similar, containing letter-style differences (such as J instead of J, the crossbar being omitted from the former) as opposed to deviations in production processes.

Overall, economy and consistency serve to reduce the potential for production errors or distortions during the handwriting process, which may consequentially interfere with the identification of a letter. The regularities in letter production processes validate a rule-based description, such as "produce strokes from left to right", or "always produce multiple vertical stokes in succession". However, we cannot always assume that there will be one optimal way of producing a given letter; there may be a number of equally efficient candidate production methods to select from. For example, a letter 'T' can be produced starting with either the horizontal crossbar or the vertical stroke, with no obvious advantage for selecting one production trajectory over the other. Letters such as these have longer onset production latencies, as well as less fluent and more variable trajectories, which are considered to be due to

conflict being generated between competing production rules (Thomassen and Tibosh 1991, Thomassen, Meulenbroek, and Tibosh 1991).

The aforementioned regularity, and potential conflict between different letter production processes has been captured in the probabilistic production-rule model of Thomassen and Tibosh (1991). This model emulates human production processes through the implementation of eight production rules, five principal and three additional, all of varying strengths. However, this model raises some outstanding questions, such as (i) how does letter production behaviour emerge through experience and practice? (ii) What are the representational characteristics of these letterforms in terms of their combined spatial and temporal properties (Viviani and Terzuolo 1982)? And (iii) how might such learning experiences influence the production of novel letterforms?

In order to address these outstanding questions we have adopted a connectionist modelling approach to explore the linearising process in letter production. Classically, connectionist models have been used to explore a variety of cognitive and perceptual tasks (McClelland, Rumelhart and the PDP Research Group 1986, Morris 1989, Elman, Bates, Johnson, Karmiloff-Smith, Parisi, and Plunkett, 1996), such as reading (Seidenberg and McClelland 1989), acquisition of the English past-tense (Rumelhart and McClelland 1986, Plunkett and Marchman, 1991, 1993, 1996, Plunkett and Juola, 1999), the development of knowledge of conservation (Shultz, 1998), and the development of knowledge of the balance scale task (McClelland 1989, Shultz, Mareschal and Schmidt 1994). Typically, these models acquire proficiency through learning without the aid of explicitly imposed rules or heuristics, with comparable performance to human data, and the added bonus of being able to explore the internal representations formed by the model during the learning

process, in order to explore the structure of the network's solution (Elman 1990, 1991).

2.   Simulating Letter Production Sequences

The aim of this investigation was to explore the emergence of letter production behaviour in a model that learns as a result of its experiences, specifically focusing on two elements (i), experience-driven behaviour – through the model's output for novel artificial letters after learning, and (ii) the impact of spatial and temporal information upon the internal representations of letters within the model. Although stroke sequence and direction have a close relationship in the process of linearising, for the purposes of this investigation, stroke sequence was deemed to be a sufficient starting point for the exploration of production behaviour. This is because it is possible to use letters that vary solely in their stroke sequences to detect production behaviour consistent with learning experience, and to explore the potential impact of both spatial and temporal information combined upon the representations of these letters.

2.1   The Task and the Model

The model's task was, given a representation of a static artificial letter at input, to identify and produce as output a sequence representing the order of the strokes used to draw the letter. The network used to carry out this task can be seen in figure 1. The first feature of this model to note is that it is recurrent, with feedback connections from the output layer to the hidden layer. These connections are necessary for a network to be able to generate a sequential output in response to a single input.

 In this model, the output layer (8 units) consisted of two parts. For both parts there was a unit corresponding to each different line-type (4 units). The task for the first part of the output layer was to simply identify the component strokes of the input

letter. The task for the second part of the output layer was to output the production sequence for these component strokes. Thus, the activations for the first part of the output layer remained constant, whilst the second part produced the sequence. The constant output on the first part of the output layer, encouraged the model to identify the component strokes of the letter, which is an essential precursor to outputting the sequence of component strokes. This type of response was included in the model's required output as it aided the initial stages of learning, and models with this form of output layer performed better than those without it.

In the example in figure 1, the letter 'L' is given as input to the model. In the first part of the output layer the two component line-types for this letter are identified (units corresponding to the vertical and horizontal line-types are shown as active). On the second part of the output layer (where the sequence of the identified component strokes is given), the unit corresponding to the vertical line-type is activated at the first time step (t_1), followed by the activation of the horizontal line-type at the second time step (t_2).It should be noted that on the second part of the output layer, only one unit is active during a single time-step.

[Insert figure 1 about here]

In order to complete the task the model needed to learn a number of sub-tasks: (i) the spatial relationship of the units in the input layer as a two-dimensional representation in a 7x7 grid (ii) to segment compound letters into their constituent line segments, and finally (iii) to produce the two output strokes in a reasonable sequence. Furthermore, the model is required to be able to carry out this task irrespective of the size or position of the letter on the input array. Of course it is (in principal) difficult to

establish exactly what the model is learning in order to accomplish this task, but the fact that the networks here are required to segment and sequence previously unseen letters suggests they really are being treated as described. Finally, this form of input focuses upon that which is required in order to perform a process, as opposed to explicitly reflecting the modality of the input (i.e. a visual stimulus). This therefore imposes the minimum necessary pre-defined structure upon the model, ensuring that the representations formed by the model were experience-driven, process-based, and self-organised.

2.2   Training the Model

The model was trained using back-propagation through time (BPTT - Rumelhart, Hinton and Williams 1986). BBPT is similar to standard back-propagation in that it operates over an activation and error back-propagation phase, with the main difference being that these phases extend over time. The selection of this training algorithm was motivated by performance considerations, given the limited capacity of Simple Recurrent Networks (an alternative choice of sequential network) to generalise (Wang, Liu, and Ahalt, 1996).

The model was trained incrementally across two phases using a cross-validation technique, where, following each epoch of training, the generalisation performance of the model was assessed using novel training exemplars. Training ceased when the model reached its best generalisation performance (lowest Mean Squared Error -MSE) for a given phase.

All letters used in this training process (which can be seen in figure 2) had production sequences consistent with naturalistic production methods, and were composed of two different line-types of varying combinations. In phase one, the

model was trained on letters one to eight, which are all consistent with a <u>left-to-right</u> production rule, and the network was trained to produce them in that order. In phase two, letters nine and ten were added to the training set (these were <u>exceptions</u> from the test set shown in figure 3). These two letters were consistent with a <u>top-to-bottom</u> production rule only. The motivation for this training regime was to explore how knowledge of these two different production processes evolved over the two phases of training.

[Insert figure 2 about here]

The exemplars used in this training process varied in terms of their size and/or their position on the input array, with component strokes ranging from three to seven units in length. The total number of size and position combinations was calculated for each letter. These were then randomly allocated to two equal groups, with one set being used to train the model on, and the other to assess generalisation performance as training progressed (each set consisted of 366 exemplars in phase one from a total of 732, and 407 exemplars in phase two, from a total of 814).

The model was trained using a learning rate 0.1, and a momentum 0.9; these values (along with the number of hidden units) were determined through systematic and incremental changes in parameter settings. A total of 15 runs were conducted in order to establish the general trend in behaviour. Each run was initialised with a different set of starting weights.

2.3   Testing the Model

The model was tested at the end of each phase of training using a set of novel letters, which can be seen in figure 3. Six exemplars of each letter, varying size and positions were used. Test letters were divided into three types: <u>normal</u>, <u>ambiguous</u> and <u>exceptions,</u> which reflect the production sequence expected from the model. For normal and exception letters, the expected production sequences were consistent with a <u>left-to-right</u> and <u>top-to-bottom</u> production sequence respectively. <u>Exception</u>s were exceptions in that after training in phase one, the model had only been trained to produce letters form <u>left-to-right</u>. For ambiguous letters, production from <u>left-to-right</u> or <u>top-to-bottom</u> would result in different stroke sequences. Thus, the normal and exceptions letters tested the model's ability to output regular production sequences (but which are described in terms of two different rules). The ambiguous letters were used to test for variations in letter production behaviour described in the literature as a result of conflict between production rules. The output sequences for each test letter according to three different production rules can be seen in table 1.

[Insert figure 3 about here]

[Insert table 1 about here]

In order to explore how the spatial and temporal knowledge of letters was represented in the model, and how this knowledge influenced the models' ability to generate stroke sequences for the novel test characters a Principal Components Analysis (PCA) of the hidden unit activations was conducted upon the model (details of this process can be found in section 3.2). The representations of letters learnt during training were probed with a subset of the training exemplars (six per letter, of varying sizes and positions, figure 4 showing three variants). This <u>probe set </u>of letters

was used, as a PCA displaying representations for all training letters would be difficult to interpret due to overcrowding. These representations were then overlaid with the representations of the test letters. A summary of the different pattern sets used in training and testing can be found in table 2.

[Insert figure 4 about here]

[Insert table 2 about here]

2.3.1   Phase One

In this phase, the model learnt to produce strokes consistent with a <u>left-to-right</u> production sequence. However, for some training letters (2, 3 and 6, as shown in figure 2) this production sequence was also consistent with a top-<u>to-bottom</u> production rule. The extent to which the model adopted the <u>left-to-right </u>production sequence for letter strokes was tested through the use of the ambiguous test letters.

 It was hypothesised that as all the letters used to train the model conformed to a <u>left-to-right</u> production sequence, that the model would attempt to produce all test letters in the same fashion, including ambiguous test letters (which would be produced from left-to-right, resulting in a horizontal-vertical production sequence for both test letters). The model was also tested on exceptions letters to explore the extent of the models' generalisation ability.

2.3.2   Phase Two

The assessment of the model during this phase was dependent upon the model's success in phase one. In phase two the model's production knowledge was extended through the introduction of letters only consistent with a <u>top-to-bottom</u>

production sequence only. This additional training permitted the representations of letters with explicitly different methods of production could then be explored. The stability of the model in terms of producing the appropriate output sequences for stable (normal) and variable (ambiguous) letters was tested.

3.   Results

All runs of the model were tested following completion of each training phase using the letters shown in figure 3. This section is divided into two sub-sections, firstly looking at the output behaviour of the model, followed by an exploration of the representations formed through the model's learning experiences, using PCA.

3.1   Output Behaviour

The output response sequences generated for each test letter were analysed as to whether they were either a <u>definite output</u> or a <u>no-response</u>. An output was considered a definite output if the activation of the target units were 0.50 or over, and no non-target units had an activation of 0.50 or higher at either time-step of the output sequence. Any other outputs were classified as a no-response. For the normal and exceptions testing subsets (see figure 3) definite outputs were compared with the appropriate output response, in order to determine their accuracy. Output response sequences for the ambiguous subset were assessed simply in terms of the sequence of production, as either stroke sequence was considered to be an appropriate output response.

3.1.1 Phase One

The ability of the model to generate definite output responses was compared with the quantity of no-responses produced by the model over all runs. An average of 64.5% responses were definite outputs. The output sequences for normal test letters was as expected with 87.4% of definite output responses produced in accordance with a left-to-right production sequence. For no-responses, outputs more often than not consisted of incomplete letter production sequences.

Interestingly, for ambiguous letters, a variety of output response sequences were generated. The production of these different sequences across these letters suggested that the model held more than one level of production description, with production sequences letters adhering to the full range of different production rules as displayed in table 1. Supporting evidence for these multiple levels of description also comes from the mixed output responses generated by the model for the exceptions subset of letters. This general trend in behaviour was shown over all runs and can be seen in figure 5. In order to illustrate in more detail the patterns of responses given by the model over runs, a selection of model runs and their response sequences are summarised in table 3. Interestingly, responses from run 3 are in accordance with to a top-to-bottom production rule, whereas responses from runs 11 and 13 are in accordance with vertical-stroke-first, and left-to-right production sequences respectively. Responses from run 7 do not correspond to a specific rule-based description of letter production processes, as was occasionally the case for some model runs.

[Insert figure 5 about here]

[Insert table 3 about here]

In summary, the analysis of the output response sequences produced by the model has found that the model is capable of producing coherent and analysable output production sequences for novel letters. Further examination of the output responses of the model indicates that the model is sensitive to the individual properties of the letters. This resulted in multiple levels of letter production descriptions being accommodated by the model, even when the model was trained on letters whose output production sequences could be described in terms of a single rule (produce strokes from <u>left-to-right</u>). This output response behaviour is likely to have occurred as a result of training the network on letters that were all produced consistently in a left-to-right manner.

3.1.2 Phase Two

The expansion of the training set in phase two to include letters that could only be described by a <u>top-to-bottom</u> production rule did not have a negative impact upon the model's overall ability to generate definite output responses (72.2% of outputs were definite responses, this increase in comparison to that of phase one, being due to the model being able to produce exception letters). There was little interference between letters from the normals test set that have different production sequences for different production-rules (62.5% of normals were produced using the appropriate production sequence, with the majority of remaining responses being indefinite output as opposed to incorrect responses). The additional training resulted in a model able to produce letters in a manner consistent with two different production rules and produce stable and consistent output behaviour within the appropriate letter context. The general pattern of the model's behaviour during this phase can be found in figure 6.

A more detailed pattern of output responses for a selection of runs can be found in table 4. As seen for phase one, production sequences for normal letters were generally found to be correct. Overall, the deviations in letter production sequence, were most prominent for letter 3 (see figure 3, and table 4). This letter is interesting because its production sequence is changed when a top-to-bottom production rule is applied. Moreover, it is a letter that is composed of diagonal line-types like the exceptions, for which the only viable production sequence is top-to-bottom. This reflects some minor interference in letter production for similar letters. It is noteworthy that this form of deviation was not found for letter 2, a letter composed of horizontal and vertical strokes, but which differs in production sequence according to the production rule used to describe it. In table 4, run 1 corresponds to a left-to-right, run 2 a vertical-stroke-first, and run 14 to a top-to-bottom production rule. Responses from run 8 do not correspond to a specific production rule over both letters. Overall, the model shows consistent production behaviour for normal and exception test letters over runs, with varied production responses for ambiguous letters.

[Insert figure 6 about here]

[Insert table 4 about here]

3.2   The Representation of Learnt Letters

The representations formed following training were explored with a view to explaining how the representations of learnt characters influenced the generalisation behaviour of the model. PCA was conducted following each phase of training, upon a subset of training patterns (the probe set: see table 2) and the generalisation set of artificial letters.   For a given set of patterns, the vector formed by the collective

activations of all fourteen hidden units were recorded at each time-step for which an output was generated (two time-steps output for each input pattern), PCA was then conducted on all the vectors of these activations. The first two principle components were then selected and the hidden unit activation vectors for both the training and test set were projected into the space spanned by these two principal components. These activation vectors can therefore be drawn as points in a two-dimensional space (Elman 1993). In this section, we first explore the representations formed by the model following training in phase one. Following this we compare the representations formed during training with those from test letters by taking the representations formed during training and overlaying them with those from testing, in order to gain an impression of which learnt representations were the most influential for a each letter in the generalisation test set

3.2.1   Training representations

The clustering of representations of letters appeared to be sensitive to general properties such as stroke-type and temporal sequence, as opposed to the grid positions in which exemplars were presented during training. Same line-types were grouped together, particularly horizontal and vertical lines (this is most likely to be due to the higher frequency of letters composed of these line-types in the training set). Temporal clustering was also observed, with distinctively separate clusters being formed for strokes produced at t_1 from strokes produced at t_2. This representational structure is shown for training patterns in figure 7, were three regions of spatial and temporal clustering and sub-clustering are highlighted. This pattern was consistent for letters composed of horizontal and vertical strokes across training phases. For letters composed of diagonal strokes, there was a tendency for some letter strokes to be

grouped in pairs (likely to be due to the additional training on *exceptions* letters in phase two).

[Insert figure 7 about here]

3.2.2 Testing representations: normal and ambiguous letters

The representations from testing with the generalisation set of letters adhered to the same representational structure as those seen in training (see figure 7). When the output production sequence for these letters was correct, the representations for these letters at t_1 were closest to those of training letters that shared the same stroke-type in the same temporal sequence. For representations at t_2, the clustering was also within line-type, but not always overlapping letters that followed the same temporal sequence. When the output was incorrect, the representations of at least one stroke were closest to a letter from training that shared the same stroke-types, but had a different temporal sequence. This is shown for *normal* letters in figure 8A, which illustrates the representations for two cases, one where the output sequence given by the model was correct, and another where the sequence was incorrect. For the correct sequence, the first stroke of the letter is in close proximity to the representations of strokes from training that is of the same type and point in the temporal sequence. For the incorrect case, the first stroke is closest to the representations from a training letter that shares the same stroke-types, but not the same temporal sequence. This is also shown to be the case for ambiguous letters – see figure 8B.

[Insert figure 8 about here]

3.2.3   Testing representations: exception letters

For exceptions, the representations from generalisation set letters were closest to training set representations of letters composed of diagonal strokes of the same type. Again, the representations of t_1 strokes from the generalisation letters for correct output sequences were closest to letters that shared the same production sequence. For incorrect outputs, the representations were inconsistent, sometimes not seeming closest to any stroke-type representation in particular. However, there were incorrect outputs for which the representations of strokes at t_1 closely matched the sequence of strokes output. The representations formed for exceptions letters are shown in figure 9.

[Insert figure 9 about here]

4.   Summary and Discussion

This paper describes a connectionist model of sequential letter production that partially undertakes the process of linearising, by producing stroke sequences for static letters presented at input. The model draws upon, and is guided by its learning experiences, as shown through its internal representations, when generating stroke sequences for novel letters, and in this respect provides an emergent account of letter production behaviour.

In terms of the pattern of output responses given by the network for different letter-types, the model generates a consistent method of producing letters that can be described in terms of a single production rule, or, have the same production sequence across different rules. For letters that have classically produced variable production sequences (which has been explained as conflict between competing production-rules:

Thomassen and Tibosh 1991, Thomassen, Meulenbroek, and Tibosh 1991), the model also showed a variety of production sequences.

In this paper, we have been careful to use the term "rules" as a means of describing the regularity of output production sequences across letters, and in this respect, as a method of characterising the regularity of output behaviour. This model contains no predefined rules prior to the onset of training. Of course, a connectionist model may embody rules of varying degrees and strengths through their connection weights (Bates and Elman, 2002). However, such rules are not absolute, and are in contrast to those explicitly defined by production systems, or, in the manner in which rules are defined in the context of the English past tense (Pinker and Ullman, 2002). Connectionist models have advocated associative similarity-based accounts for behaviour (cf. Kinder and Assman 2002, McClelland and Patterson 2002, Plunkett and Marchman 1991, Rumelhart and McClelland 1986). Whether this model contains rules or not would be the focus of further investigation. However, it can be said that this model exhibits similarity-based regularities in letter production, and indeed the distinction between rules and such regularities may not be the dichotomy it appears to be (Pothos, 2005).

The internal representations of the model show how the model draws upon its learning experiences in order to generalise in the production of the novel test letters, with those that share the same stroke types and temporal sequence being the most similar, clustering together. Strokes of the same letter commonly grouped together, and in this respect the identity of a given letter was preserved through the relative similarity of the representations. These representations highlight two important properties of letters for production: (i) the spatial properties of the letter, and (ii) the temporal component to production. In a production task these two aspects of a letter

are co-dependent, and trajectory-specific (Richardson, Davey, Peters, Done, and Anthony 2002). In the model, the grouping of stroke-types (the spatial aspect) was very prominent in the representations formed, however, for a letter to be produced correctly the representations (particularly those at t_1) were required to be in a point in representational space that fitted both the spatial and temporal properties of the letter stroke.

Finally, over the learning process, the model was introduced (in phase two) to new and different production information regarding some characters. Interestingly, the model was capable of assimilating this new and different information, which influenced the model's production behaviour for letters that had similar properties, but did not have a marked impact upon the model's ability to generate appropriate letter production sequences. This learning did affect the representations of these letters within the model, which highlights that the model was able to detect where the new information was relevant, and adapt accordingly in order to assimilate the new knowledge, but also at the same time maintain the commonalities for other letters (for which the new knowledge was not relevant), indicating that the knowledge acquired by the model was within context, and sensitive to the stroke-specific properties of the letter - both spatial and temporal (Viviani and Terzuolo 1982). These representations may offer some insight into the structure and organisation of higher-level motor programs for letter production (Teulings and Shoemaker 1993, van Galen and Webber 1998).

In conclusion, this model adds emergent perspective to the letter production process, in that both regularity and variability for letters of differing properties may evolve through a process of learning, and in this respect complements the probabilistic production model of Thomassen and Tibosh (1991). In our model, the relative

similarity between learnt and novel letters drives production behaviour. Generally, whether the letter production system evolves into a system of rules of varying strengths or not, this model shows how letter production may emerge as a result of experience and practice.

References:

Alston, J., and Taylor, J., 1987, *Handwriting: theory, research and practice* (New York: Nichols).

Alston, J., and Taylor, J., 1988, *The handwriting file* (Cambridge, UK: Living and Learning Ltd).

Alston, J., and Taylor, J., 1990, *Handwriting Helpline* (Manchester, UK: Dextral Books).

Armstrong, S., 1994, *Handwriting* (London: The English and Media Centre).

Bates, E., and Elman, J., 2002, Connectionism and the study of change. In *Brain development and cognition: A Reader (2nd edition)*. Edited by M. Johnson (Oxford, UK: Blackwell Publishers).

Desbiez, D., Vinter, A., and. Meulenbroek, R., 1996a, The influence of task features on biomechanical exploitation in drawing. In: *Handwriting and drawing research: Basic and applied issues*, edited by M. Simner and C. Leedham (Amsterdam, The Netherlands: IOS Press) pp. 57-70.

Desbiez, D., Vinter, A.., and Meulenbroek, R., 1996b, Biomechanical and perceptual determinants of drawing angles. **Acta Psychologica 94**, 253-271.

Elman, J.L., 1990, Finding structure in time. **Cognitive Science 14,** 179-211.

Elman, J.L., 1991, Distributed representations, simple recurrent networks, and grammatical structure. **Machine Learning 7,** 195-255.

Elman, J.L., 1993, Learning and development in neural networks: the importance of starting small. **Cognition 48**, 71-99.

Kinder, A., and Assmann, A., 2002, Learning artificial grammars: No evidence for the acquisition of rules. **Memory and Cognition 28,** (8), 1321-1332.

Marquardt, C., Gentz, N., and Mai, M., 1996, On the role of vision in skilled handwriting. In: *Handwriting and drawing research: Basic and applied issues*, edited by M. Simner and C. Leedham (Amsterdam, The Netherlands: IOS Press) pp. 87-97.

McClelland, J., and Patterson, K., 2002, Rules or connections in past-tense inflections: What does the evidence rule out? **Trends in Cognitive Sciences 6,** (11), 465-472.

McClelland, J., Rumelhart, D., and The PDP Research Group (Eds.). 1986, *Parallel Distributed Processing: Volume 2. Psychological and biological models* (Cambridge, MA: MIT Press).

Meulenbroek, R., and Thomassen, A., 1991, Stroke-direction preferences in drawing and handwriting. **Human Movement Science 10**, 247-270.

Meulenbroek, R.., and Thomassen, A., 1993, Exploitation of elasticity as a biomechanical property in the production of graphic stroke sequences. **Acta Psychologica 82**, 313-327.

Meulenbroek, R., Thomassen, A., Schillings, J., and Rosenbaum, D., 1996, Synergies and sequencing in copying L-shaped patterns. In *Handwriting and drawing research: Basic and applied issues*, edited by M. Simner and C. Leedham (Amsterdam, The Netherlands: IOS Press) pp. 57-70.

Morris, R. 1989. (Ed.). *Parallel Distributed Processing: Implications for psychology and neurobiology* (New York: Oxford University Press).

Pinker, S., and Ullman, M., 2002, The past and future of the past tense. **Trends in Cognitive Sciences 6,** (11), 456-463.

Plunkett, K., and Marchman, V., 1991, U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. **Cognition 38,** 1-60.

Plunkett, K., and Marchman, V., 1993, From rote learning to system building: Acquiring verb morphology in children and connectionist nets. **Cognition 48,** 21-69.

Plunket, K., and Marchman, V., 1996, Learning from a connectionist model of the English past tense. **Cognition 61,** 299-308.

Pothos, E , 2005, The Rules versus Similarity Distinction. **Behavioural and Brain Sciences 28,** 1-49.

Richardson, F., Davey, N., Peters, L., Done, J., and Anthony, S. 2002, Connectionist models investigating the representations formed in the sequential generation of characters. ESANN Proceedings of the 10th European Symposium on Artificial

Neural Networks, 24-26 April (Evere, Belgium: D-side publications), pp .83-88.

Rumelhart, D., Hinton, G., and Williams, R.. 1986, Learning internal representations by error propagation. In *Parallel Distributed Processing: Volume 1*, by D, Rumelhart, J.McClelland, and the PDP Research Group. *Parallel Distributed* (Cambridge, MA: MIT Press) pp. 318-364.

Rumelhart, D, and McClelland, J., 1986, On learning the past tense of English verbs. In J. McClelland and D. Rumelhart, and the PDP Research Group (Eds.), *Parallel Distributed Processing: Volume 2. Psychological and biological models* (Cambridge, MA: MIT Press). pp216-271.

Seidenberg, M., and McClelland, J., (1989), A distributed, developmental model of word recognition and naming. **Psychological Review 96,** 452-477.

Teulings, H., and Shomaker, L., 1993, Inariant properties between stroke features in handwriting. **Acta Psychologica 100,** 275-290.

Thomassen, A., and Tibosch, H., 1991. A quantitative model of graphic production. In J.Requin & G.E. Stelmach (Eds.), *Tutorials in motor neuroscience* (Dordecht, Netherlands: Kluwer).pp 269-281.

Thomassen, A., Meulenbroek, R., and Tibosch, H., 1991, Latencies and kinematics reflect graphic production rules. **Human Movement Science 10**, 271-289.

Van Doorn, R., and Keuss, P., 1993, Invariant properties between stroke features in handwriting. **Acta Psychologica 82,** 275-290.

Van Galen, G., and Weber, J. 1998, Online size control in handwriting demonstrates the continuous nature of motor programs. **Acta Psychologica 100,** 195-216.

Viviani, P., and Terzuolo, C., 1982, Trajectory determines movement dynamics. **Neuroscience 7**, 431-437.

Wang, D., Liu, X., and Ahalt, S. 1996. On temporal generalisation of simple recurrent fnetworks. **Neural Networks 9,** 1099-1118.

Figure 1

Figure 2

Figure 3

Normal: produced left-to-right

1   2*   3*   4

Ambiguous: variable sequence of production

5   6

Exceptions: produced from top-to-bottom

7   8
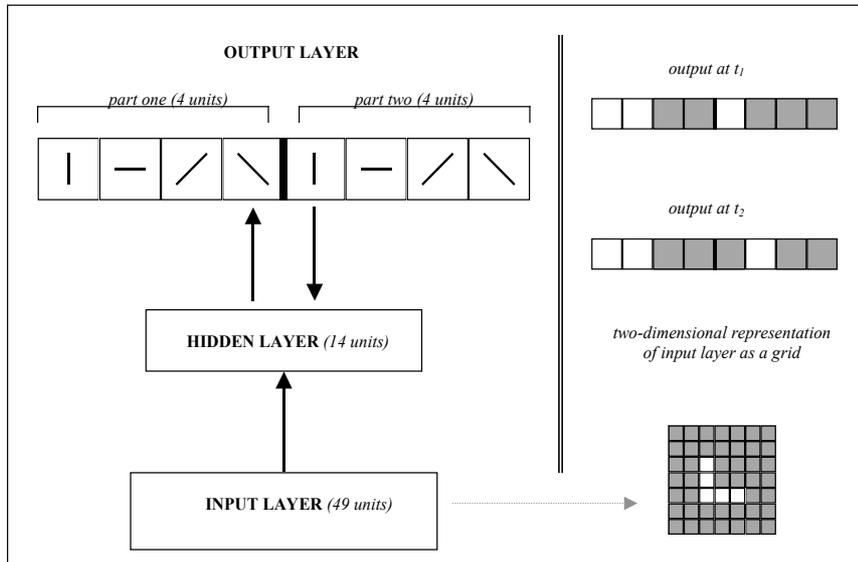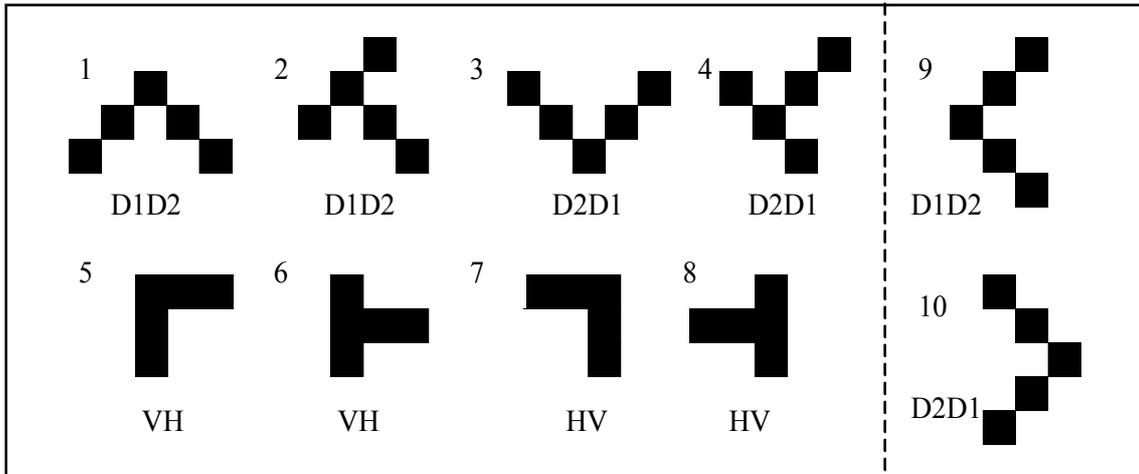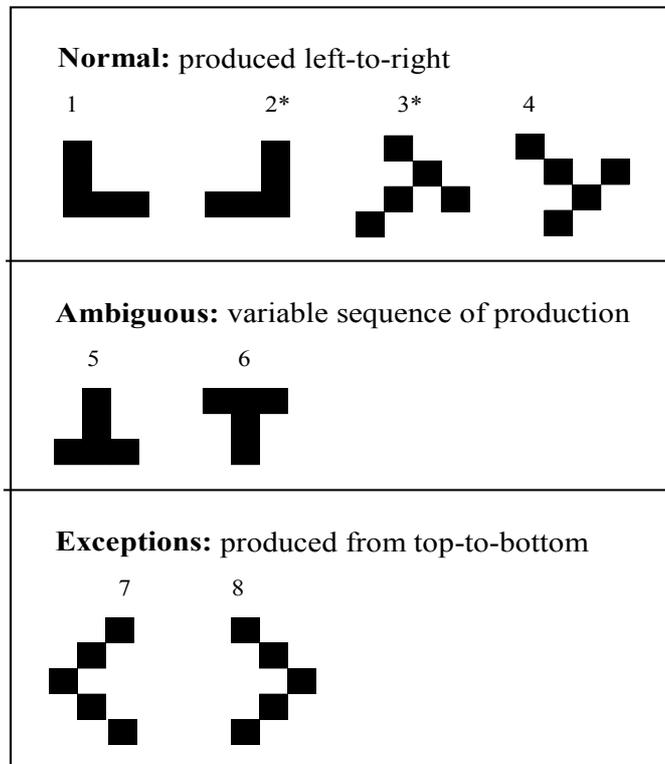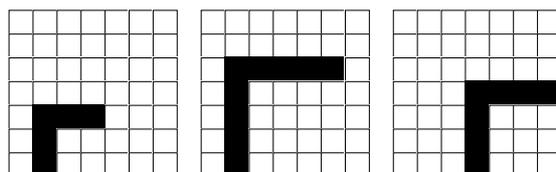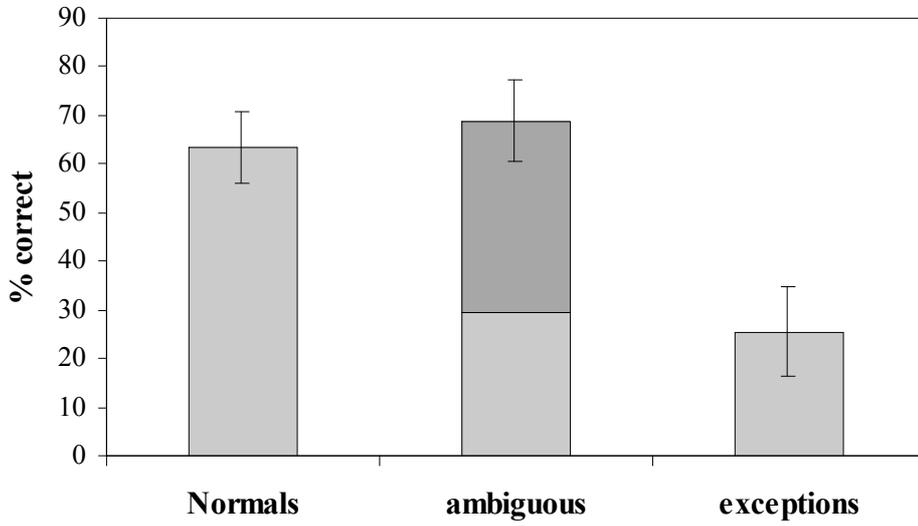
Figure 4

Figure 5

Figure 6
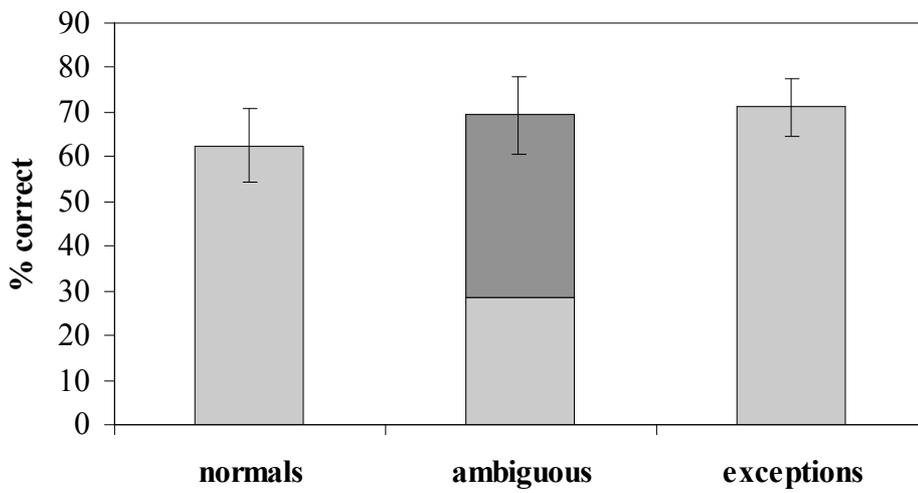
Figure 7
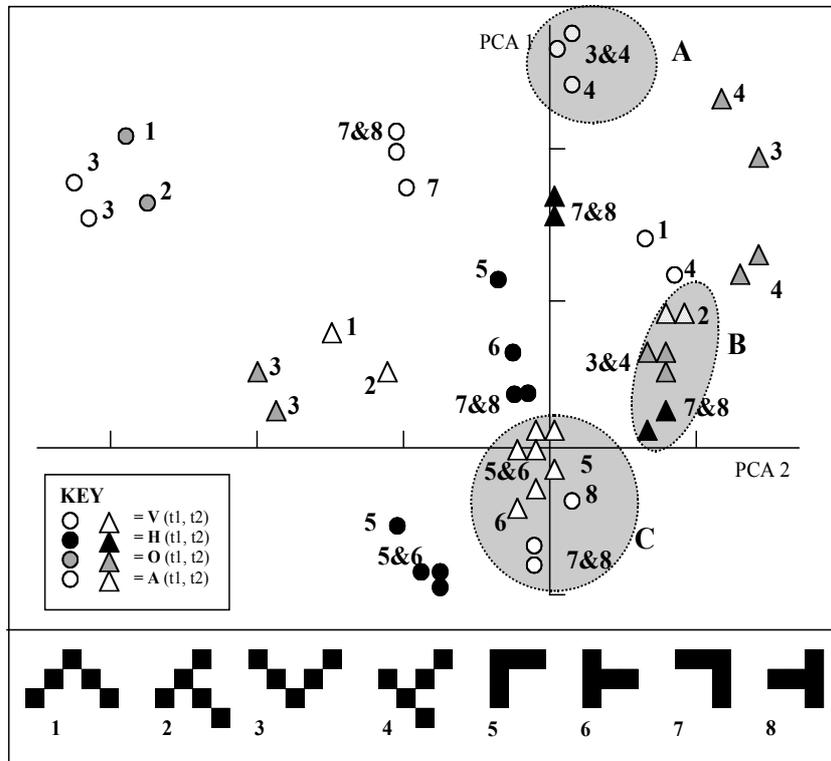
Figure 8
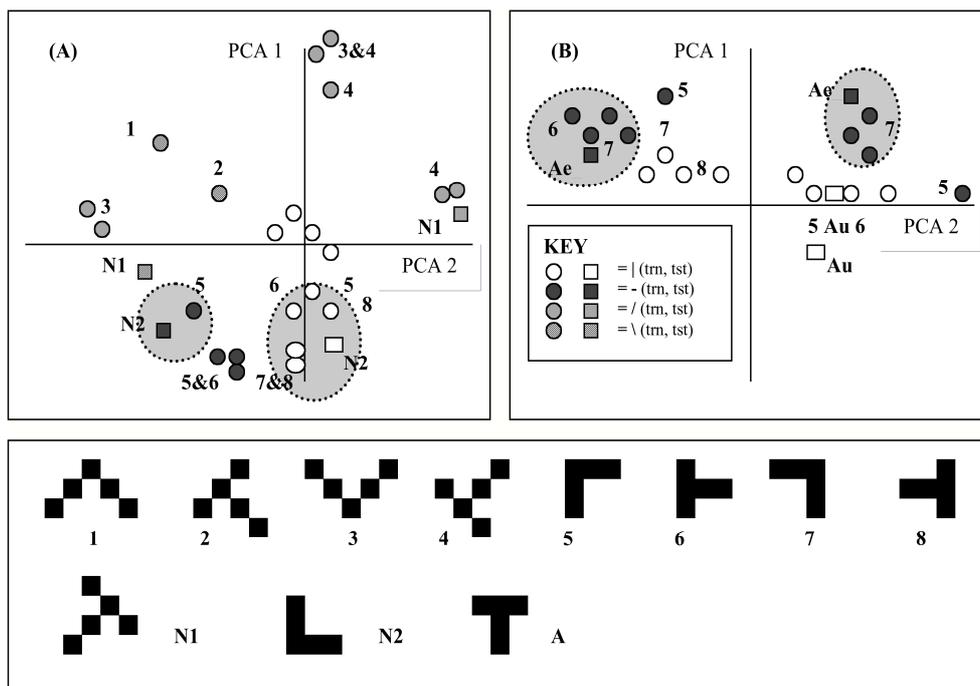
Figure 9
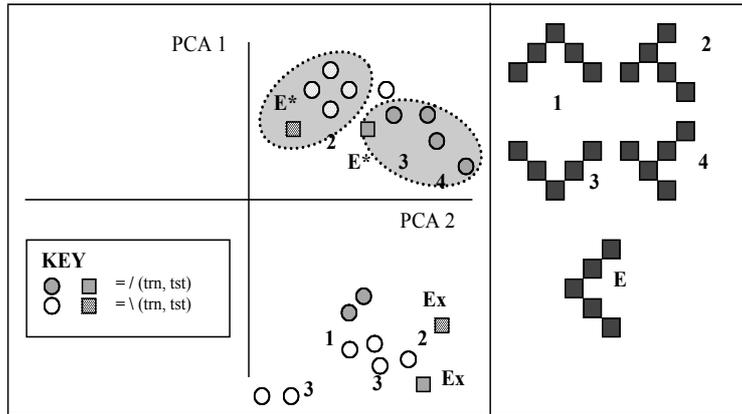
Table 1

	Normals				Ambiguous		Exceptions	
	1	2	3	4	5	6	7	8
Left-to-right	|_	_|	/\	\/	_|	_|	-	-
Top-to-bottom	|_	|_	\/	\/	|_	_|	/\	\/
Vertical-stroke-first	|_	|_	-	-	|_	|_	-	-

Table 2

Name of Pattern Set	Number of Exemplars	
	Phase One	Phase Two
Training Set	366	407
Generalisation Set	366	407
Test Set	48	36
Probe Set	24	30

Table 3

	Normals				Ambiguous		Exceptions	
	1	2	3	4	5	6	7	8
Run 3	*	*	*	*	|_	_|	o	o
Run 7	*	*	*	*	_|	|_	*	x
Run 11	*	*	*	*	|_	|_	x	x
Run 13	*	*	*	*	_|	_|	o	x

Table 4

	Normals				Ambiguous		Exceptions	
	1	2	3	4	5	6	7	8
Run 1	*	*	o	*	_|	_|	*	*
Run 2	*	*	*	*	|_	|_	*	*
Run 8	*	*	*	*	_|	|_	*	*
Run 14	*	*	*	*	|_	_|	*	*

Figure 1. A network performing the sequential production task. Active units are displayed in black, with arrows indicating full connectivity between layers. The network is presented with a representation of a static (artificial) letter as input. As output the network is required to identify the component line-types of the letter (as displayed in the static section of the output layer) and also output the sequence of strokes that should be used to produce it (on the sequential section of the output layer).

Figure 2. Shows the artificial letters presented to the network during training. Letters one to eight were presented in phase one. Letters nine and ten were added to the training set in phase two. Line-type abbreviations showing the production sequence for each artificial letter are also indicated (Vertical, Horizontal, Diagonals, D1 / and D2 \).

Figure 3: The test set of novel letters, composed of line-type combinations familiar to trained networks. The test set is divided into three subsets: normals, ambiguous and exceptions, with each subset testing a different aspect of letter production behaviour. For letters marked with a '*', the implementation of a different production rule would result in a different stroke sequence.

Figure 4: An example of a set of three probe input patterns from the training set of letters. Each letter is a different size and location upon the input array.

Figure 5: The mean percentage of correct outputs (with error bars showing the standard deviations) for each test set over all runs of the model for phase one. Outputs for the ambiguous subset are shown for HV (bottom) and VH (top) ordered combinations. It can be seen that over runs, the model was able to produce a large number of correct/appropriate output responses for normal and ambiguous letters, but struggled with exceptions.

Figure 6. The mean percentage of correct outputs for each test set over all network runs for phase two (with error bars showing the standard deviations). Correct outputs for the *ambiguous* set are shown for HV (bottom) and VH (top) ordered combinations. It can be seen that networks were able to produce a large number of correct/appropriate output responses for all test sets, thus exhibiting a rule-like behaviour for normal and exception test letters, and varied output responses for ambiguous letters.

Figure 7: Shows the results of a PCA displayed as a 2-D graph from a randomly selected network from phase one. The first two principal components account for 59% of the total variance. There is a distinctive clustering of same line-types, especially among those that are produced at the same time-step in a sequence. For example region A shows clustering of same line-types, which share the same temporal sequence. Region B shows clustering of lines that share the same temporal sequence, with sub-clustering of same line-types. Region C shows clustering of the same line-type, with two sub-clusters, one corresponding to t_1 and the other to t_2.

Figure 8: PCA's from randomly selected runs (A) shows representations formed for two *normal* letters. The output sequence for letter N1 was incorrect, where that for N2 was correct. (B) shows representations formed for ambiguous letters for an ambiguous character, for Ae, an expected sequence at phase one (HV), and Au, an unexpected sequence (VH). These two principal components account for 59% and 55% of the total variance respectively. The representations of strokes for correct test letters are close to those of training letters, which share the same stroke types (as indicated by the highlighted regions in map space).

Figure 9: Shows the representations for an exceptions letter, as a result of PCA. The first two principal components account for 60% of the total variance. As indicated by the highlighted regions in map space, the representations of strokes for the correct test letter (E*) are close to corresponding stroke representations of training letters. This is not the case for Ex, for which an incorrect output sequence was produced.

Table 1: Shows the stroke sequences for each test letter according to three different production rules (shown as strokes, read from left to right).

Table 2: Shows a summary of the pattern sets used in both phases.

Table 3: Shows a summary of the pattern of letter production sequences for the full test set of letters (labelled 1 to 8, as in Figure 3) for four runs of the model in phase one, with responses summarised over the majority of definite outputs. * = correct response, x = incorrect response, o = no consistent response sequence. The responses for ambiguous letters shown as strokes (read from left to right).