Four letters good, six letters better: Exploring the exterior letters effect with a split architecture.

John Hicks, Jon Oberlander & Richard Shillcock
Division of Informatics
University of Edinburgh
Scotland
United Kingdom
(All correspondence to: John.Hicks@ed.ac.uk)

Abstract

Recent models employing split neural networks have demonstrated that such architectures are effective for processing visual information. Furthermore, it has been shown that certain emergent strategies of processing are particular to these split architectures. We investigate one such strategy, the exterior letters effect, extending and generalizing it, and go on to discuss the implications that effects which are marked in split architectures bring to bear on lateralization and hemispheric specialization in human cognition.

Introduction

What might be the advantages for bi-hemispheric processing of visual information? How does real-time high-density information management—such as that employed in the human visual system—cope with the fact that processing of the same thing is done in two halves, in two different places? What is it about the interaction between the hemispheres that allows for the apparently automatic co-operation between them? The answers to these central questions inform almost all other areas of cognition, and discussion of them abounds in the literature. And yet modeling studies on such aspects of gross brain morphology remain relatively under-developed, in spite of the nervous system's clear division centralised in two cerebral hemispheres. The complex relationship that comes into play between particular architectural features and general processing strategies, as well as distinct variations in the nature of the stimuli involved, can play a large role in empirical studies. Although clearly the techniques implicit in learning and execution of a task could be multifarious, models such as the one presented here assist in teasing apart the details of dual processing. Split-architecture connectionist models of cerebral function take as their motivation the well known psychology of the hemispheres, but open out onto a field that is largely uncharted.

Background

When cognitive science per se was still in its infancy, studies on split brain phenomena were well underway (Gazzaniga, 1970). Work with patients who had undergone commissurotomy made it clear that the two halves of the brain could function autonomously when disconnected. The highlight of this discovery was the apparent inability of the right hemisphere to speak for itself in any real sense (Gazzaniga, 1983). Thus, a century after its initial stipulation, Broca's hypothesis gained even more secure footing. At the same time, the disparate activity resulting from two hemispheres out of touch with each other, and, in particular, the speechless fumblings of the right-side, gave a real sense to the distance neuro-anatomically (and thus perhaps experientially) that lay between the hemispheres. This was a distance that was unbridgeable through subcortical structures in the event that the corpus callosum was cut (although see (Sergent, 1987)).

Such severe unlinking is by no means the only evidence of separate identity of the hemispheres. The visual field is split vertically about the fovea in the retina, the right and left halves of the visual field projecting contralaterally into the cortical regions of the left and right hemispheres respectively (Sperry, 1968; Fendrich & Gazzaniga, 1989). Because of this, large scale degradations which are specific to one hemisphere, can lead to marked behavior in tasks reliant on apprehension of the entire visual field, as in cases of unilateral neglect. This deficit, afflicting right-hemisphere stroke victims, manifests itself commonly in the line-bisection task (Halligan & Marshall, 1998; Reuter-Lorenz & Posner, 1990), where the affected portion of the visual field is essentially omitted by subjects asked to designate the midpoint of a line.

The clear contralateral routing of information to opposite hemispheres by the visual system affords a lot of ground for research in normals as well. Key issues about general pattern recognition, symmetry and particularly face recognition can be addressed (Bruce, Cowey, Ellis, Perrett, 1992). Similarly, work in word recognition (e.g. Rumelhart & McClelland , 1981) must at some level be affected by the constraints of the visual processor; assuming the gaze is focussed around the midline of a word, interactionist accounts of processing have to deal at least with the transference of visual information to the locus of letter activation, if not simultaneous activation in different hemispheres.

Jordan's account of letter activation (1990, 1995) bears on the current study. With subjects focusing on a fixation point, stimuli of 200msec or less, containing letter strings of a fixed length but without a full complement of letters (e.g. "d__k" is a two letter string of length four) were presented and masked with a null string of identical length. Subjects were asked to report the letters that appeared. Significantly, letters coming at the edge points of the string length were more robustly reported than letters that came from interior positions. This "exterior letters effect" (ELE) forms the vehicle for the current discussion on split architectures, and has already been successfully replicated in a connectionist network using a divided "visual field" (Shillcock and Monaghan, in press) each side of which projects separately to one of two hidden layers.

Reggia, Goodall, & Shkuro (1998) describe a word read-

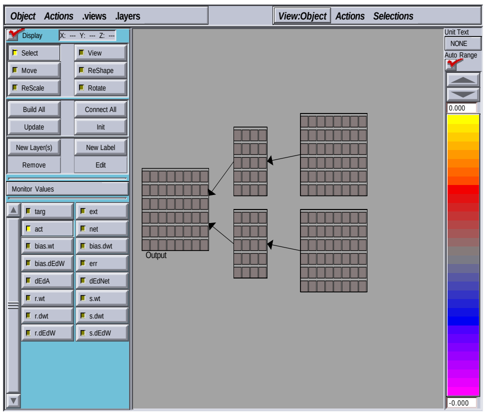

Figure 1: A typical instantiation of a split architecture network, shown here with the aid of the PDP++ graphical interface.

ing task which is learned by a split network. The task is a vehicle for gauging the effects different network parameters have on the degree of lateralization in the fully trained net, lateralization being determined by a "winner take all" competition between two hidden layers given a single input layer. Other modeling work on lateralization deals with the nature of the respective topographies, in terms of cortical organization (Alvarez, & Reggia, 1998; Levitan & Reggia, in press), while elsewhere Shestova & Reggia (1999) do relate a visual identification task to which our models bears an implicit resemblance, insofar as there is a "dual route" strategy for the reception of input.

Qualitative data using split network

Shillcock and Monaghan (in press) describe a network in which the input field and the hidden units are split in two. With a network similar to that pictured in Figure 1, they present lexical input to the network, but include a positioning technique which allows the four letter words to move across the visual field, being presented in any one of five positions (from occupying only the left hemi-field to occupying only the right hemifield, passing through the midpoint, where two letters of the four are projected to each side, halfway). It is at root this method of data presentation that ensures that the split net can and will develop a strategy for solution that is not found in the non-split control.

This effect, which relates to Jordan's work as described above, manifests itself as a diminished reliance in the trained network on the interior letters of words, with a related robustness for recognition for letters in word-final and word-initial positions. Sucn networks seem to exploit the exterior letters to a greater extent than the nonsplit networks. We claim that the preferential treatment of the exterior letters is provoked by the manner of presentation and the current study is intended to expand upon this idea.

To sum up Shillcock and Monaghan's findings: there is an ELE, comparable to that found with human subjects, demonstrated by their model. After training the networks with a split architecture showed a significant advantage in recognition of the exterior letters when degraded stimuli consisting of the original words with either the interior or exterior letter pair "masked" with an ambiguous activation pattern. This finding was true in their study for all positions across the two visual fields. The study was slightly limited however; only four-letter words were used. These are a special case, containing two interior and two exterior letters. Below we explore the affect in the six letter case, also expanding on the criteria used to measure the effect.

Modeling with a Split Architecture

Rarely are claims made that align connectionist models directly with cellular components of the cortex, upon which the design and operation of simulated neural nets may nevertheless be based.

This caveat is even more salient within the split architecture paradigm, each of the hidden layers ostensibly standing for an entire hemisphere to which input is projected. Other things being equal, it is important to avoid such direct correlations between the neural level and the grain of the model.

Experiments

Two experiments were performed. For each one, a number of different simulations were run using split and non-split network designs. Each simulation was repeated 10 times and the results all reflect averages for the 10 runs. Subsequent tests using degraded stimuli employed each of the 10 trained nets for that class, the results again being averaged. Details of the nets and the stimuli are given below.

Materials

A series of simulated neural networks, employing a back-propagation learning algorithm, was trained using the top 60[1] four and six letter words of English respectively. Also used was a list of 60 random strings of the same length[2]. The words were coded following the system of Plaut and Shallice (1994), assigning 8-bit features to each letter, each feature representing an aspect of letter orthography such as "contains closed area" etc. The coded words were then presented to the network through a shift invariant identity mapping (SIIM) task which maintains the integrity of the stimulus organization, while moving it sequentially along the input window. Input nodes that fall outside the location of the word at any time have activation zero, as do the inactive bits within the eight bit feature vector of each letter. The vertical split in the input reflects that of the fovea and thus, as a word is repeatedly presented to the network from all possible positions across the input, it crosses from one "visual hemifield" to the other, activation being redirected to the associated hidden layer accordingly.

Separate networks were used for the four and the six letter tasks, but the number of hidden units, 20, remained the same in each case. Nets not possessing a split hidden layer were used for a control task in which a simple visual field (containing the same number of input units as the non-split network). Networks featured full feed forward connectivity between the layers, save in the case of the non-split models,

[1]Ranking of the words was based on frequency counts from the celex lexical database.

[2]The distribution of letters in these strings was absolutely flat, in opposition to the skewed frequency counts for high frequency words of English.

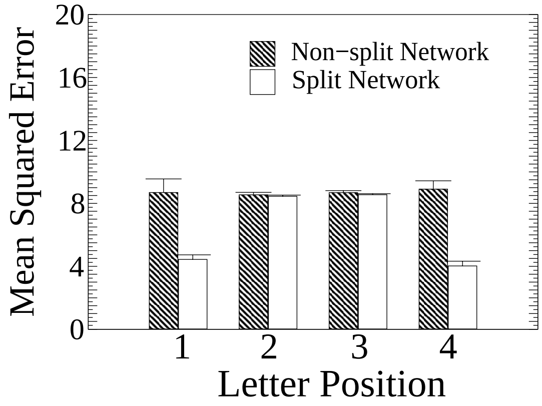

Figure 2: Comparison of nonsplit and split models when networks trained to recognize a set of random strings are fed degraded stimuli, where only the exterior letters are present. The error for the exteriors in the split model is much diminished.

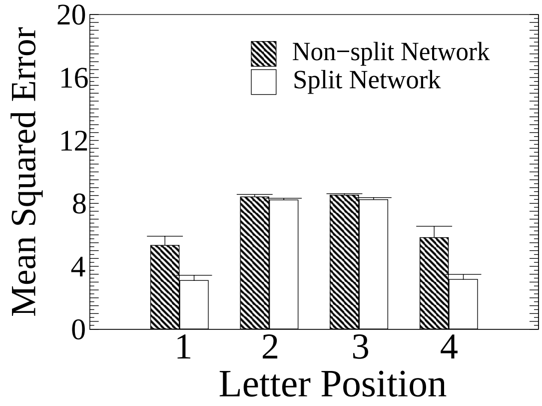

Figure 3: Comparison of nonsplit and split models when networks trained to recognize the English word set are fed degraded stimuli, where only the exterior letters are present, and the interiors masked.

in which the connectivity between the input and hidden layers underwent a random pruning of half of the connection. This was to ensure that the network's power was consistent with its split counterparts, network power being directly proportional to number of weighted connections.

For all simulations the PDP++ Neural Nets software from CMU was used, running on an Ultra 5 work station.

Results

Experiment One: Replication of previous Results

In attempting to replicate the exterior letter effect that Shillcock and Monaghan showed, we trained split and non-split networks on the English and non-word stimuli. As their simulations mirror Jordan's recognition task for exterior letters, and this involved the presentation of degraded or masked letter strings to trained nets, we used a similar technique. However, it is worth pointing out that we also found a general advantage in *word* recognition for the split networks. This, of course, relates to the size and nature of the lexicon and overall error at the output layer, whereas the letter recognition task is defined in terms of individual letter positions.

On the individual letter scores, for stimuli in which the interior letters were rendered ambiguous, Shillcock and Monaghan found an effect similar to Jordan's empirical finding, namely that recognition of exterior letters was favorable in such conditions, but significantly more so when the network employed a split architecture. This preference is seen in Figure 2 for non-words and Figure 3 for words. Paired t-tests (two-tailed) checking relative error of exterior letters across networks ($df = 19$) gave $t = 14.73, p < .0005$ for the study in non-word strings and $t = 23.32, p < .0005$ for that involving English words, a highly significant effect representing an advantage for the split net in both cases.

Rather than the specific presentation of degraded stimulus that Shillcock and Monaghan demonstrate, to generalize the effects of the split architecture, if they are indeed robust, a more general technique is helpful. The effects of masking letter pairs in strings becomes inordinately complex with strings even of 6 letters, as we later found (five types of masking means at least a 10 way comparison of masked *on each network*) and generally, we would like to find a more all encompassing and straightforward view of network behavior, in terms of letter position error after training, for example. To this end we compared the two models, without using masked words.

However, although we were able to replicate and even generalize Shillcock and Monaghan's findings to a degree, by using degraded stimuli, we found that the effect itself did not significantly cross over into analysis of error levels by letter position as a whole, as Figure 4 shows.

Experiment Two: Extension of ELE

In the second experiment, our attention was directed to the networks' performance on the learning task with the six letter stimuli. Again, training consisted of learning over all positions in the visual field, with two different stimulus sets; the top 60 six letter words of English and 60 pseudo words, or random letter strings.

While in the case of the four letter stimuli no significant difference could be demonstrated using error by letter position, for the six letter case there was indeed a notable difference in network performance as seen below. Figure 5 shows the error for each letter position after non-split and split networks had both been trained on the non-word stimuli. In this case a fairly significant drop in error was registered. Taking the difference in error between exterior letters and their adjacent interior letters, we then compared the differences in these (i.e. has the network error dropped significantly for one of the networks on the exterior letters?). $df = 9$ for each of the following two tailed t-tests: the word initial pair, for each network, $t = 6.64, p < .0005$; the word initial pair in the split network compared with the word final pair in the non-split network, $t = -3.47, p < .007$; the word final pair in the split network compared with the word initial pair in the non-split network $t = -2.49, p < .034$; and the word final pair for both networks $t = 4.65, p < .001$. These figures in the main corroborate the story told by the graph: that the split network

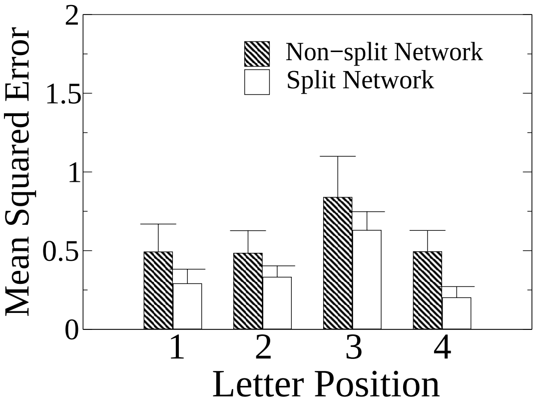

Figure 4: Comparison of nonsplit and split models for the top 60 4 letter words of English. The error is registered after 400 training epochs. Although the error drops across the board for the split model, it does so uniformly, the exterior letters showing no advantage (the best result from 4 separate interior-exterior letter error comparisons between different networks architectures, using a two tailed t-tests, $df = 9$, gave $t = -2.89, p < .018$)

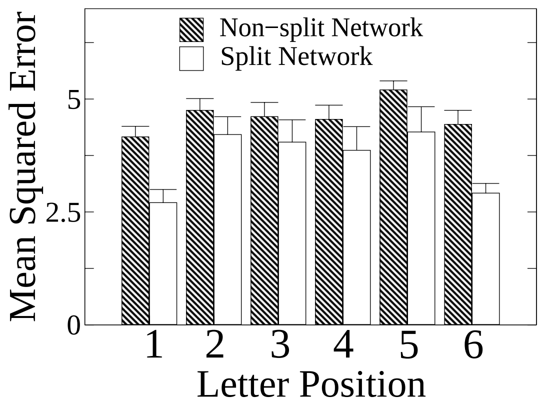

Figure 5: Comparison of nonsplit and split models for 60 random strings of 6 letters each. The error is registered after 400 training epochs. See text for details.

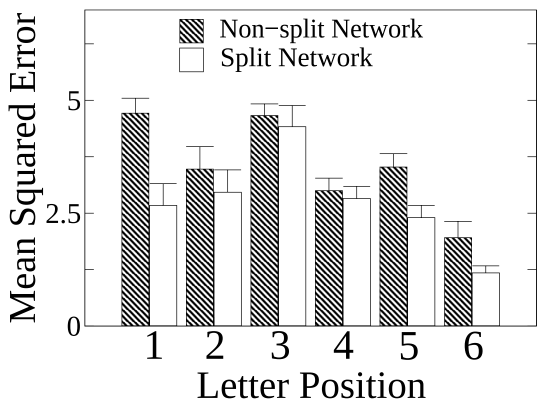

Figure 6: Comparison of nonsplit and split models for the top 60 6 letter words of English. The error is registered after 400 training epochs. See text for details.

purchases more success using outside letters than the non-split network. This is statistically clearest for the first and last of the above comparisons, where the only difference was the network architecture (cross word comparisons, e.g. word initial with word final, admit interference from the stimuli). A similar comparison within each network (i.e. seeing if there was a significant drop in performance between interior pairs and exterior pairs not linked to a change in network architecture) yielded, $t = .97, p < .359$, for the non-split net, $t = .54, p < .603$, for the split, or, no difference.

Figure 6 shows the results for the different nets after training with the English word stimuli. As above, $df = 9$ for each of the following two tailed t-tests: the word initial pair, for each network, $t = 6.30, p < .0005$; the word initial pair in the split network compared with the word final pair in the non-split network, $t = -12.07, p < .0005$; the word final pair in the split network compared with the word initial pair in the non-split network $t = -6.81, p < .0005$; and the word final pair for both networks $t = 2.84, p < .019$ The significant dip in the error of exterior letters reiterates the trend shown in the graph. Of particular interest here is the form of the "arch" in the error by position of the split network, as well as the quasi-sinusoidal effect the non-split net seems to find when presented with the English word strings. These topics are taken up in the general discussion.

Discussion

In this study we have performed experiments with a series of split and non-split neural networks. The results re-affirm the main finding of Shillcock and Monaghan, that a difference in network performance is based on the architecture, split or non-split, that that network employs. Shillcock and Monaghan's model produced an ELE, which says exterior letters of strings are favored in conditions of stimulus degradation. This effect was demonstrated by them under very specific conditions, which we were able to generalize as holding across the board for degraded stimulus[3]. The effect, a large drop in relative error by the split network for exterior letters only, is clearly seen in the corresponding figures (2 and 3). We tried but were not able to extend Shillcock and Monaghan's results still further using simple error monitoring criteria, whereas with six letter strings the simple error metric not only revealed the ELE, but did so strongly.

In general, the ELE can be seen as the benefit of having a split hidden layer. With a single hidden layer, the mapping learned by the network for each pattern at each letter position is highly interdependent. Thus instantiations of letters at one position are much more likely to be conflated with their immediate neighbors. What a separate layer for each visual field buys is a foothold for representational independence. The same mapping is learned in either case, but with

[3]It is worth noting that in their actual task, Shillcock and Monaghan read error at single presentation positions of input, as well as the corresponding letter position at output; that is, although they examined every position, we demonstrated the cumulative effect of the error at different positions.

the split network, error back-propagated from the output to hidden layers during learning brings each hidden layer into line with the other through an indirect coordination. Thus a modicum of independence in each layer is retained, and this is used as collateral against an investment, or specialization, that that layer makes in direct proportion to the input it is exposed to. And this input favors, in the case of each hidden layer, the exterior letters of the stimulus, by simple fact of relative exposure (interior letters disappearing across the "fovea" and into the other hemi-field sooner for every pattern presented). This potential "separateness" for marginal phenomena (i.e. exterior letters) licenses, amongst other things, the robust behavior in the face of degraded input the split network demonstrated.

Other questions remain, however. For example, although for the six letter case we were able to show preferential learning for exterior letters just by monitoring error by letter position, the four letter case yielded no such view. A possible reason for this is network competence in terms of the capacity of the hidden layer to find a secure mapping from input to output. The total number of hidden units was the same in both nets; yet the six letter strings required not only a larger input area (two visual hemi-fields of *six*, as opposed to four, letters each), but they also constituted a much larger input set in general, as each word appeared in each possible position (five for the four letter model, but *seven* for the six letter case). Thus at the lower end of the extreme, the smaller net manages its quarry rather elegantly, the residual shape of the error by letter (Figure 4) probably reflecting nothing other than the structural regularities present in English orthography. When this competence envelope is pushed, as in the case of increasing the task load on the hidden units with the introduction of a six letter mapping, the hidden layer is forced to resort to economic measures, visible as the ELE. Indeed, this would provide some explanation of why, at the four letter level, the ELE can only be detected with finer method, the presentation of corrupted input.

If the effect is a conflation of these two trends, then that goes some way to explaining the "arch" of the split bars in figure 6: the pressure on the net to retain as much as it can means a sacrificing of the representations of interior letters in favor of the exterior representations which are easier for each hidden layer to maintain. The contrary shape of the nonsplit network in the same figure suggests that it needs to resort to a different strategy[4] in order to degrade gracefully under the increased weight of the six letter task.

These suggestions form but a part of a larger set of topics to which modeling with a split architectures gives rise. There are many others besides, not the least of which is a retention of the intuitive notion known as "modularity" *at some level* in the brain. At one time, connectionist models threatened to rule out the idea of "separate parts" altogether. The current study is one which demonstrates the integration of two concepts: the benefits brought by separation–e.g. the independence of the hidden layers as a means of exploiting presentational regularities, like exposure to exterior letters, which are themselves brought about *for free* through the relationship that obtains between the model and the environment; and the importance of concentration, as that of the units within the hidden layers, without which the error driven learning of such problems would not be possible at all.

[4]Perhaps one not unlike taking English C-V-C phonological regularities, or rather the way they manifest in orthography, as a template for resolving input.

Acknowledgments

The first author acknowledges the generous support of the EPSRC in this work.

References

Alvarez, S, & Reggia, J. (1998). Metrics for Cortical Map Organization and Lateralization. *Bulletin of Mathematical Biology, 60*, 27–47.

Bruce, Vicki (Ed)., Cowey, A. (Ed)., Ellis, Andrew W. (Ed)., Perrett, D. I. (Ed). (1992). *Processing the facial image*. Clarendon Press/Oxford University Press, Oxford, England.

Fendrich, R. & Gazzaniga, M.S. (1989). Evidence of foveal splitting in a commissurotomy patient. *Neuropsychologia*, *27:3* 273–281.

Gazzaniga, M.S.(1970). *The Bisected brain*. Appleton-Century-Crofts; New York.

Gazzaniga, M.S. (1983). Right hemisphere function following brain bisection: A 20 year perspective. *Am. Psychol.*, *38* 525-549.

Halligan, Peter W. & Marshall, John C. (1998). Visuo-Spatial neglect: The ultimate deconstruction? *Brain & Cognition*, *37:3* 419-438.

Jordan, T.R. (1990). Presenting words without interior letters: Superiority over single letters and influence of postmark boundaries. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 893–909.

Jordan, T.R. (1995). Perceiving exterior letters of words: Differential influences of letter-fragment and non-letter-fragment masks. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 512–530

Levitan, S & Reggia, J. (in press). Interhemispheric Effects on Map Organization Following Simulated Cortical Lesions. *AI in Medicine*.

Plaut, D. C. & Shallice, T. (1994). *Connectionist modeling in cognitive neuropsychology: A case study*. Lawrence Erlbaum Associates, Inc; Hove, England.

Reggia, J, Goodall, S, & Shkuro, Y. (1998). Computational studies of lateralization of phoneme sequence generation. *Neural Computation*, *10*, 1277–1297.

Reuter-Lorenz, Patricia A & Posner, Michael I. (1990). Components of neglect from right-hemisphere damage: An analysis of line bisection. *Neuropsychologia, 28:4* 327-333.

Rumelhart, D. E. & McClelland, J. L. (1981). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, *89*, 60–94.

Sergent, J. (1987). A New Look at the Human Split Brain. *Brain, 110*, 1375–1392.

Shestova, N & Reggia, J. (1999). A Neural Network Model of Lateralization during Letter Identification. *Journal of Cognitive Neuroscience*, *11:2*, 1277–1297.

Shillcock, R, and Monaghan, P. (in press). The computational exploration of visual word recognition in a split model. *Neural Computation*.

Sperry, R. W. (1968). Apposition of visual half-fields after section of neocortical commissures. *Anatomical Record*, *160* 498–499