Evolving Processing Speed Asymmetries and Hemispheric Interactions in a Neural Network Model

Alexander Grushin and James A. Reggia

{agrushin,reggia}@cs.umd.edu

Department of Computer Science, University of Maryland, College Park, MD 20742

Abstract: Substantial experimental data suggests that the cerebral hemispheres have different processing speeds, and that this may contribute to hemispheric specialization. Here, we use evolutionary computation models to examine whether asymmetric hemispheric processing speeds and lateralization can emerge in neural networks from the need to respond quickly to stimuli and/or to minimize energy consumption. Simulated neuroevolution produced networks with left-right asymmetric processing speeds whenever fitness depended on energy minimization, but not on quickness of response. The results also provide support for a recent hypothesis that subcortical cross-midline interactions are inhibitory/competitive.
Keywords: Lateralization; Asymmetry; Corpus callosum; Neuroevolution; Genetic algorithm

1. Introduction

Cerebral hemispheric specializations, where one hemisphere performs a task better than the other, include language, handedness, and visuospatial processing, among others [3]. The underlying causes of hemispheric specialization/lateralization are not well understood at present. The many known underlying neurobiological asymmetries are presumed to be relevant, but which of these actually contribute to hemispheric specialization, and the role of hemispheric interactions via the corpus callosum, remain important unanswered questions. Given the above uncertainties, a substantial number of neurocomputational models of emergent cerebral specialization and hemispheric interactions have been created over the last fifteen years (reviewed in [9]). Typically, such neural models consist of corresponding left and right cerebral regions, often connected to each other via a simulated corpus callosum, that undergo a learning period involving alterations in synaptic strengths, during which lateralization emerges due to underlying asymmetries. These past studies have generally only examined how hemispheric specialization arises from a pre-existing asymmetry, but not how such asymmetries arose during evolution.

In the research described here, we extend a recent computational study [11], addressing a different, currently unresolved question: Why might the presence of underlying hemispheric asymmetries be beneficial to the brain from an evolutionary point of view [7]? We evolve the parameters in a genetic representation of interacting left and right hemispheric regions to identify conditions under which underlying asymmetries will emerge from an evolutionary process, and subsequently during learning/development lead to behavioral specialization. We extend the earlier work [11] by using evolutionary multiobjective optimization methods [1] to examine the specific hypotheses: 1. hemispheric regions will evolve to have asymmetrical processing speeds if they must both learn to perform well during a developmental period, and at the same time, minimize energy utilization and/or response time, two factors that have previously been postulated to play important roles in brain evolution and the emergence of lateralization [2,4,10]; and 2. inhibitory cross-midline influences will also evolve under these conditions [8].

2. Methods

The neural networks employed in this study (Figure 1) have a very simplified architecture motivated by top-level brain structure. For any given network instance, input patterns activate the left and right subcortical regions, which compute their activation in an iterative fashion, with both regions iterating simultaneously, although perhaps at separate frequencies. After one unit of time, the subcortical regions stop iterating, and pass their results to the cortical regions, which iterate for another time unit, and then activate the output region, which iterates either until each output neuron is on or off within a tolerance of 0.2 (at least one output neuron is required to be on, since all activation levels are initially zero), or for a maximum of three time units. The kth iteration of region X with a frequency of F_X iterations per time unit (where $1 \leq k \leq F_X$) is performed at time k/F_X, where $X \in \{LS, RS, LC, RC, OUT\}$ in Figure 1.

A single iteration consists of two phases. In the first phase, each neuron i in a region computes a linearly-weighted sum h_i of its incoming synaptic activity. The second phase consists of updating the activation levels a_i as follows: $a_i \leftarrow a_i + \delta[\sigma(h_i) - a_i]$, where $\sigma(x) = (1+e^{-x})^{-1}$ is the logistic function, and $\delta = 0.1$ is the time step. Feedforward weights are initially drawn from a uniform distribution in [-1, 1] using a random number generator,
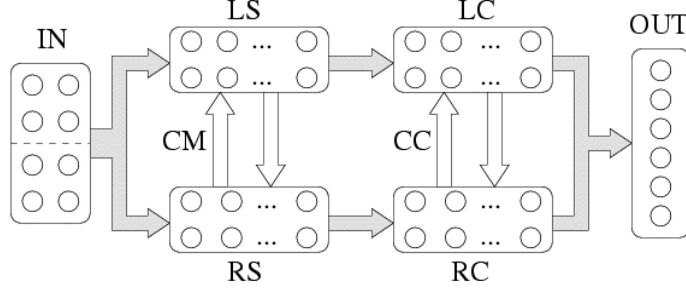
Figure 1: Network architecture. The regions are: input (IN), left/right subcortical (LS/RS), left/right cortical (LC/RC), and output (OUT). Dark arrows denote full feedforward connections, while light arrows denote subcortical cross-midline (CM) and corpus callosum (CC) connections, which are strictly pairwise.

and are subsequently trained via error backpropagation. Training is done for 300 epochs in a per-pattern fashion, with a learning rate of 0.2.

The uniform weights W_{CM}, W_{CC} of interhemispheric connections and the iteration frequencies F_X of regions do not change during training, but are instead determined over a larger time scale by a genetic algorithm using tournament selection. This algorithm runs for 50 generations with a population size of 50, a tournament size of 3, and with crossover, mutation, and reproduction-only probabilities of 0.5, 0.4, and 0.1, respectively. The evolved parameters (Table 1; S is the random number generator seed) are encoded as a sequence of real numbers ("genes") in the range [0, 1], and realized via a linear mapping. Because of the enormous computational cost of repeatedly training a population of 50 neural networks during the evolutionary process (a typical evolutionary run took in the order of $15 - 20$ hours of CPU time on a 1 GHz Pentium III processor), we used very small networks: $|IN| = 8$; $|LS| = |RS| = 10$; $|LC| = |RC| = 15$; and $|OUT| = 6$.

For every generation of the evolutionary process, the fitness of each chromosome is evaluated by instantiating a network with phenotypic parameter values that are based on its genes, and measuring its post-training performance on a problem (taken from [11]) that requires a mapping of 16 input vectors in $\{0, 1\}^8$ into 16 output vectors in $\{0, 1\}^6$. The performance metrics are described in Table 2, where P is the set of pattern pairs; d_{pi} and o_{pi} are the final desired and observed activation values, respectively, for neuron i, given pattern p; t_p is the number of time units necessary for the output neurons to satisfy the termination conditions for pattern p; and F_X^{max} is the maximum possible iteration frequency of region X (20 in all cases). Since there are multiple possible metrics for estimating energy consumption in a biological brain [11], two alternate definitions were

3

Table 1: Evolvable Attributes for the Network

Position in the Chromosome	Parameter	Minimum Value	Maximum Value
1	S	0	10000000
2	F_{LS}	1	20
3	W_{CM}	-3.0	3.0
4	F_{RS}	1	20
5	F_{LC}	1	20
6	W_{CC}	-3.0	3.0
7	F_{RC}	1	20

Table 2: Performance Metrics ("Costs")

Metric	Formula
E (error)	$E = \left(1/|P| \cdot \Sigma_{p \in P}[\Sigma_{i \in OUT} (d_{pi} - o_{pi})^2] \right)^{-1/2}$
t_r (response time)	$t_r = 1/P \cdot \Sigma_{p \in P}[\ t_p\ ]$
J_{act} (activity-based energy consumption)	$J_{act} = \Sigma_{p \in P}[A_p / A_p^{max}]$, where $A_p = \Sigma_{X \in \{LS,RS,LC,RC\}}[|X|^2 \Sigma_{i \in X}\ o_{pi}^2]$, and $A_p^{max} = \Sigma_{X \in \{LS,RS,LC,RC\}}|X|^3$
J_{freq} (frequency-based energy consumption)	$J_{freq} = \Sigma_{p \in P}[(B_p)^2 / (B_p^{max})^2]$, where $B_p = \Sigma_{X \in \{LS,RS,LC,RC\}}[|X| \cdot F_X]$, and $B_p^{max} = \Sigma_{X \in \{LS,RS,LC,RC\}}[|X| \cdot F_X^{max}]$

used: J_{act}, which is based on activation levels within the subcortical/cortical regions, and J_{freq}, which depends on the iteration frequencies of those regions. From these cost metrics, four objectives (fitness criteria) to be maximized were defined as follows: accuracy $\equiv 2 - E$; speed $\equiv 4 - t_r$ (this is a behavioral definition, and does not refer to the underlying hemispheric processing speeds/frequencies); activity-based energy efficiency $\equiv 2 - J_{act}$; and frequency-based energy efficiency $\equiv 2 - J_{freq}$. In the six experiments reported here, six different fitness functions were used: I. accuracy alone; II. accuracy and speed; III. accuracy and activity-based energy efficiency; IV. accuracy and frequency-based energy efficiency; V. accuracy, speed, and activity-based energy efficiency; and VI. accuracy, speed, and frequency-based energy efficiency. For multiobjective fitness functions (II – VI above), fitness during evolution is determined by the *Strength Pareto Evolutionary Algorithm* (SPEA) [12], an evolutionary multiobjective optimization technique [1]. Rather than keeping one current best solution, as is done in the single-objective case, SPEA maintains a set of solutions (of size ten in our experiments) that are not dominated by any other solutions in the current population,

4

promoting greater population diversity. A solution x is said to dominate another solution y if and only if no objective is satisfied by solution y better than it is satisfied by solution x, and there exists at least one objective that is satisfied better by solution x than by solution y. The fitness of a given solution depends on what solutions it dominates and what solutions it is dominated by, and is assigned in a way that rewards non-dominated solutions, and promotes a uniform sampling of different objective tradeoffs [12].

Ten independent runs of the genetic algorithm were performed with each fitness function I – VI. Each run had an initial population that was generated with a distinct stream of random numbers, but used an identical procedure otherwise. For the ten best networks returned under I and the 100 non-dominated networks produced for each fitness function II, …, VI, the post-training values of the metrics of Table 2 were recorded, along with evolved structural parameters. Furthermore, to determine the degree of lateralization, the error E_L (E_R) of each such network was measured with only the left (right) hemisphere contributing to the output. Early termination was not used during lateralization measurement; rather, each time a network was activated on pattern p, its output region was made to iterate the exact same number of times that it iterated when both sides were contributing. The lateralization coefficient ρ [5] used in this study was computed as $\rho = |\ [E_L–E_R]\ /\ [2 - (E_L+E_R)/2]\ |$. The asymmetry coefficient α was also calculated as follows: $\alpha = s_\rho \cdot |\ [F_{LS} + F_{LC} - F_{RS} - F_{RC}]\ /\ [F_{LS} + F_{LC} + F_{RS} + F_{RC}]\ |$, where $s_\rho = 1$ if the asymmetry is "in the direction" of the lateralization, and -1 otherwise. Informally, a negative value of α indicates that the lower-frequency regions are more accurate.

3. Results

We first describe the evolution of asymmetries and lateralization, and then the evolution of cross-midline subcortical and callosal connections. Table 3 gives the mean (μ) and standard deviation (σ) measures computed over all networks (one per run for I and ten per run for II – VI) produced by each of ten runs for each fitness function. Pre-training error values were generally well above 1.0, with no early termination ($t_r = 3$). Consistent with [11], the use of accuracy alone in fitness criterion I did not produce strong asymmetries or lateralization, although mild asymmetries were present, suggesting

5

Table 3: Experimental Measurements for Fitness Functions I – VI

Measure	I		II		III		IV		V		VI	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
E	0.30	0.00	0.31	0.01	0.76	0.26	0.71	0.29	0.55	0.23	0.51	0.23
t_r	1.56	0.17	1.25	0.23	2.86	0.25	2.75	0.41	2.41	0.59	2.36	0.59
J_{act}	0.25	0.03	0.25	0.03	0.09	0.04	0.12	0.06	0.14	0.05	0.17	0.06
J_{freq}	0.83	0.10	0.91	0.07	0.21	0.14	0.18	0.19	0.43	0.29	0.35	0.22
W_{CM}	-1.65	1.48	-2.36	0.74	-1.34	1.94	-1.18	1.83	-1.54	1.65	-2.38	0.83
W_{CC}	0.98	1.81	0.99	2.13	-2.10	1.25	-0.35	2.19	-2.24	1.10	0.82	2.01
ρ	0.13	0.12	0.11	0.06	0.43	0.21	0.42	0.24	0.42	0.26	0.44	0.25
α	0.03	0.06	0.01	0.04	0.77	0.19	0.62	0.29	0.53	0.36	0.43	0.33

that perfect synchronization in the iteration of opposite regions is not necessarily optimal for good accuracy. Each hemisphere performed very poorly by itself, indicating that the network's good post-training performance relies on interhemispheric cooperation.

Fitness function II, based on both accuracy and speed (quickness of response) also did not result in strong asymmetry and lateralization. In fact, the networks tended to be even more symmetric, with little correlation between underlying asymmetry in update frequency and lateralization, which is indicated by the fact that α is very close to zero. Iteration frequencies were typically high in all regions (about 19 on average), reflecting the fact that both accuracy and speed of response can benefit from high processing speeds. Therefore, E and t_r did not represent major tradeoffs, but were complementary, even if small-scale tradeoffs (where a slight sacrifice in one criterion yields a slight improvement in the other) did exist.

In contrast to the above, strong asymmetries and lateralization arose whenever either energy measure was incorporated into the fitness function (III – VI). This occurred because the model could achieve a reasonable level of accuracy while saving on energy by reducing the computational resources of one hemisphere and relying heavily on the other (more so than by moderately reducing resource consumption in both hemispheres). As shown in Figure 2, the values of asymmetry α and lateralization ρ measures are especially high for networks where a "compromise" approach between accuracy and energy efficiency is taken by the optimization process. In such cases, there exists a balance between error and energy consumption, which are negatively correlated among the evolved networks, and thus represent a tradeoff. For example, fitness based on
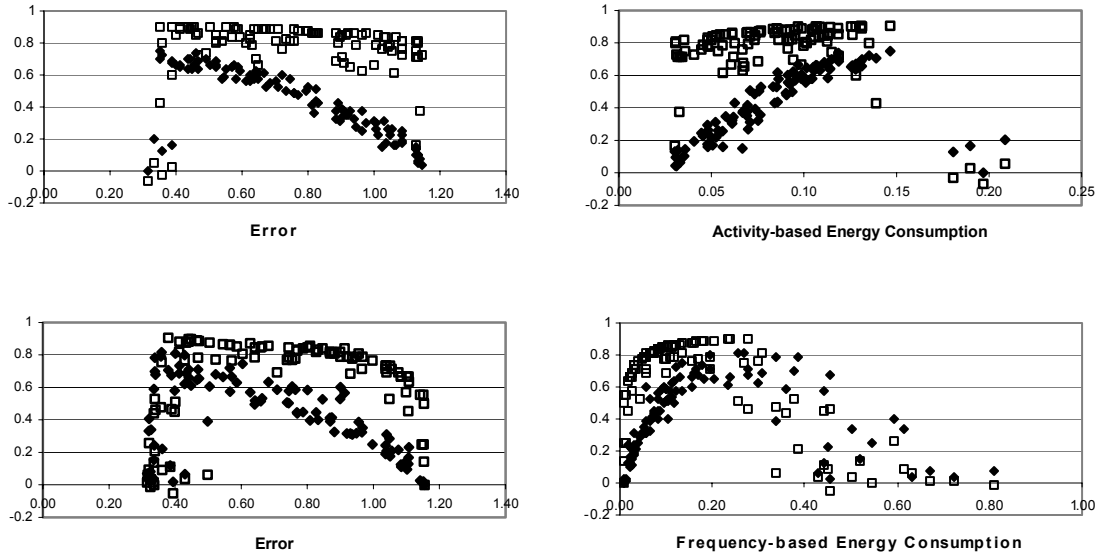
Figure 2: Scatterplots for fitness functions III (top) and IV (bottom), showing asymmetry α (□) and lateralization ρ (♦) versus error (left), and energy use (right).

accuracy and activation-based energy use (III) yielded highest values of α and ρ in the energy use range of [0.10 – 0.15], which corresponds to error values that are low, but not minimal among the evolved networks. In this compromise range, *extreme* asymmetries can occur in individual networks, where the subcortical/cortical regions have an update frequency of 20 on one side, and 1 on the other. Asymmetry and lateralization values become progressively smaller as this compromise is shifted in favor of energy efficiency, because the dominant hemisphere tends to become slower as well. In the higher-energy range, fitness function III produced four distinct networks (appearing as a separate cluster) with both hemispheres updating at a high frequency, but IV tended to create a wider variety of non-dominant hemisphere frequencies, resulting in a greater diversity of α-values.

The results for the fitness functions based on three objectives (V, VI) are, for the most part, not drastically different from the corresponding results produced under III and IV. Once again, greatest asymmetry and lateralization values are observed for networks in the moderate energy consumption range, which balances with moderate-to-high accuracy and moderate-to-low speed levels, the latter criteria being positively correlated, and therefore, complementary. Like the use of the frequency-based definition of energy consumption (as opposed to the activity-based definition), the incorporation of an evolutionary

7

pressure for speed seems to have a diminishing effect on the average asymmetry evolved (as shown in Table 3) and a strengthening effect on the diversity of networks in the higher-energy range; these effects are therefore strongest for VI.

Finally, we consider the evolution of the uniformly-weighted synaptic strengths of subcortical cross-midline connections (W_{CM}) and callosal connections (W_{CC}), as shown in Table 3. In all cases, there was a tendency for inhibitory subcortical cross-midline connections to evolve ($W_{CM} < 0$). In contrast, callosal connections were excitatory ($W_{CC} > 0$) whenever energy use was not a factor in fitness (I, II), but were usually inhibitory when energy minimization was a factor (III – V). Activity-based energy was economized by maintaining left-right inhibitory connections that allowed interhemispheric communication without excessive cortical excitation, which is the main contributor to this energy consumption measure. In the case of fitness function IV (using J_{freq}), the mean value of W_{CC} is much less negative than it is for III (using J_{act}), since the frequency-based energy measure does not directly depend on activation levels within the network. Interestingly, fitness function VI was the only one that produced strong lateralization with underlying asymmetries, inhibitory subcortical cross-midline connections, and an excitatory corpus callosum, a set of results argued to be most consistent with empirical data [8].

4. Discussion

Our neurocomputational model is drastically simplified when compared to biological brain structures, capturing only the elementary left-right brain architecture at a high level and the basic nature of neural activation dynamics and synaptic plasticity. Thus, the results we obtained can only be viewed as preliminary, and require confirmation and further study with larger and more realistic models. Still, our results, as summarized in Table 4, are intriguing, and suggest a number of conclusions.

First, when paired left and right neural pathways evolved with fitness based on both accuracy and quickness of response, significant left-right asymmetry in hemispheric processing speeds and subsequent functional lateralization did not arise. This result does not provide support for the hypothesis that rapid responses to stimuli are an important factor in the emergence of brain asymmetries related to hemispheric processing speeds

Table 4: Summary of Results

Fitness Function	Metrics in Fitness Function				Results		
	E	t_r	J_{act}	J_{freq}	α and ρ	W_{CM}	W_{CC}
I	✓				low	−	+
II	✓	✓			low	−	+
III	✓		✓		high	−	−
IV	✓			✓	high	−	−
V	✓	✓	✓		high	−	−
VI	✓	✓		✓	high	−	+

(e.g., [10]), but can be explained by the aforementioned complementary relationship that exists between speed of response and accuracy within our model. Second, and in contrast, whenever minimization of energy consumption was a factor in determining fitness, substantial asymmetry in left versus right processing speeds appeared, and led to strong functional lateralization. This result is consistent both with substantial experimental data about asymmetrical hemispheric processing speeds [6] and with arguments that the minimization of energy use is an important factor in brain evolution [2,4]. Thus, our results suggest that tradeoffs between objectives, and not just objectives themselves, produce asymmetries and lateralization. Finally, in line with the earlier results of [11], in all simulations, regardless of which factors were used in the fitness function, simulated subcortical cross-midline connections evolved to be inhibitory, while simulated callosal connections became either excitatory or inhibitory. These results are consistent with a recent hypothesis that subcortical cross-midline inhibitory connections or other subcortical left-right competitive mechanisms may be much more important in causing hemispheric specialization than is generally recognized [8]. However, they do not shed light on the ongoing controversy about the nature of callosal connections, an important problem that merits further investigation [8].

Acknowledgements

References

[1] Coello, C. (2002). Evolutionary Multi-Objective Optimization: A Critical Review.

Evolutionary Optimization, R. Sarkar et al. (eds.), Kluwer, 117-146.

[2] Gibbons, A. (1998). Solving the Brain's Energy Crisis. *Science*, 280, 1345-1346.

[3] Hellige, J. (1993). *Hemispheric Asymmetry*, Harvard University Press.

[4] Laughlin, S. & Sejnowski, T. (2003). Communication in Neuronal Networks. *Science*, 301, 1870-1874.

[5] Lezak, M. (1995). *Neurological Assessment*, Oxford University Press.

[6] Nicholls, M. (1996). Temporal Processing Asymmetries Between the Cerebral Hemispheres: Evidence and Implications. *Laterality*, 1, 97-137.

[7] Provins, K. (1997). Handedness and Speech. *Psych. Review*, 104, 554-571.

[8] Reggia, J., Goodall, S., Shkuro, Y., & Glezer, M. (2001). The Callosal Dilemma. *Neurological Research*, 23, 465-471.

[9] Reggia, J. & Levitan, S. (2003). Hemispheric Interactions and Specialization. *Handbook of Brain Theory and Neural Networks*, M. Arbib (ed.), MIT Press, 525-528.

[10] Rogers, L. (2000). Evolution of Hemispheric Specialization: Advantages and Disadvantages. *Brain and Language*, 73, 236-253.

[11] Shkuro, Y. & Reggia, J. (2003). Cost Minimization During Simulated Evolution of Paired Neural Networks. *Cog. Sys. Res.*, 4, 365-383.

[12] Zitzler, E. & Thiele, L. (1999). Multiobjective Evolutionary Algorithms. *IEEE Transactions on Evolutionary Computation*, 3, 257-271.