“Modeling visual lateralization and interhemispheric communication”

Ben Cipollini

Introduction

Lateralization is intertwined with virtually every function that we think makes us human, including language, fine motor skills, and visuo-spatial processing. In my PhD thesis, I attempted to give a computational account of lateralization in visual processing, offering hypotheses about its neural underpinnings, neurodevelopmental origins, and relationship to interhemispheric transfer. To do so, I performed new analyses on existing models, used them to account for new data, and extended them to model developmental processes. The result is a computational model of lateralization in visual processing, new models and ideas about interhemispheric transfer across species (including humans), and a number of quantitative and qualitative neuroanatomical predictions.

The thesis is separated into three main projects, each focused on data from a different discipline. In the first project (Chapters 2-4), I used neural networks to model a key type of connection in visual processing, long-range lateral connections in cortex. In Chapters 2 and 3, I examined how variations in the anatomy of these connections leads to spatial frequency biases in the model, which lead to processing differences that mirror those found in behavioral experiments using visual hemifield presentation of hierarchical letters, faces, spatial frequency gratings. Finally, in Chapter 4, I showed that well-known developmental constraints can cause this connection asymmetry in the model.

These hemispheric models were not connected. In preparation for modeling hemispheric interactions between the hemispheres, I followed up with two projects, each examining influential papers about how anatomical and physiological properties of the corpus callosum affect interhemispheric interactions.

The first of these projects (Chapter 5) focused on modeling results that suggest that larger communication delays across callosal connections weaken interhemispheric communication (Ringo, Doty, Demeter, & Simard, 1994). I re-analyzed the neural network results, and found that delay magnitude does not change interhemispheric communication - it only delays it! I then used a similar neural network to show that conduction delay variability can decrease interhemispheric communication, and argued that this variability is present early in development due to biophysical properties of immature connections.

The final project (Chapter 6) focused on comparing intrahemispheric and interhemispheric connection fiber counts across large- and small-brained species, using allometric regression (Rilling & Insel, 1999), a technique common to biological anthropology. The original paper concluded that the size of the corpus callosum is decreasing on an evolutionary time scale. I expected that their conclusion was wrong, and that there would be no decrease with a more accurate analysis. Through a careful analysis of the data in the literature, introducing several novel corrections to the data, I found to my surprise that they underestimated the reduction. In understanding the significance of this large reduction, I discovered a new way to think about the data in terms of functional connections, which I think is more relevant to the brain’s processing. I showed that this new measure explains Rilling & Insel's results, and scales perfectly (with a slope of 1) across mammals. I concluded that the relative size of the functional connectivity of the corpus callosum is not changing with brain size.
Part I: Modeling lateralization in vision

The left and right hemispheres differ in their ability to process visual stimuli that contain essential information in different frequency bands. These include stimuli containing information at low frequencies (faces), high frequencies (words), or both (hierarchical letter stimuli, spatial frequency gratings). The most influential current model of these differences simply assumes the spatial frequency modules are built into the model to account for behavioral differences (Ivry & Robertson, 1998). However, the Ivry & Robertson model does not have a story about the neural mechanisms for these frequency channels, nor how they would arise developmentally.

In my thesis, I took a completely different approach, based on previous work in the lab. I wondered how much of the data could be explained by modeling an anatomical difference between the hemispheres, rather than a functional difference. I started from the model architecture introduced by Hsiao, Shahbazi, and Cottrell (2008) and focused on a hypothesized asymmetry in long-range lateral connections. These connections are found within the grey matter and selectively interconnect nearby “patches”. In retinotopic areas such as primary visual cortex, patches have distinct receptive fields; the connections allow the patches to share information with each other. This helps boost local computations within the patch when the stimulus is weak (Levitt & Lund, 2002) or when task-based expectations about the stimulus are strong (Swadlow & Alonso, 2009). These connections are also highly active when stimuli contain familiar contours or shape information—both cases in which a stimulus feature’s bottom-up processing can be enhanced to improve accuracy and response time. These are all cases that tend to show greater lateralization in behavioral studies (see Sergent (1985) and Christman (1989) for reviews).

Long-range lateral connections are a likely involved in lateralization of vision.

The model I adopted is a special type of autoencoder (Hsiao et al., 2008) called the “differential encoding” network. Autoencoders encode inputs by learning to reproduce them at the output through an intermediate layer of hidden units. By so doing, the hidden unit’s activation becomes an alternate encoding of the pixels, a mix of features shared across images. In this autoencoder, each hidden unit has a spatial position relative to the input. Rather than connecting to all of the input units, each hidden unit randomly samples a small number of inputs to connect to. Each hidden unit samples the same number of connections, but the standard deviation of the sampled Gaussian distribution differs between hemispheres. The connections to the output are the same as those sampled to the input (see Figure 1), so this models the interactions between those nodes. In order to make this model interpretable as a model of long-range lateral connections, I used parameter values (number of connections, spread of the connections) estimated from the literature and

---

1Presented at COGSCI 2012 and 2014 (best perception/action modeling paper), VSS 2013
2The data and analysis code for all of these experiments is available at http://github.com/guruucsd/DifferentialEncoding.
Figure 2: This explains how I used the “differential encoding” neural network to model long-range lateral connections. The number of connections per hidden unit was 12, a number consistent with the literature for “patchy” connections (Amir et al., 1993) (only 6 are shown here, for visual clarity). The number of hidden units equaled the number of pixels in the input image (850), with one hidden unit positioned at each input pixel location (only one of 850 is shown here, for visual clarity). Each hidden unit connected with the input pixel at its position, in addition to 11 nearby pixels sampled from a Gaussian distribution. This allowed the hidden unit activations to simulate the activation that that input location would receive from itself (activity persists over time) and its neighbors (through the lateral connection network).

To test hemispheric asymmetry, in each simulation I created two networks—one with connections sampled from a narrow Gaussian, and one from a wide Gaussian, but otherwise identical. Based on a finding in auditory cortex and the preliminary results of Hsiao et al. (2008), I expected the narrow network to show learning / processing biases like the right hemisphere, and the wide network to show learning / processing biases like the left hemisphere.

The model shows the hypothesized frequency processing differences, without being programmed to do so.

I trained the two networks on a set of images. I then examined at what spatial frequency bands each network more accurately learned compared with the other. As predicted, I found that the network with more dense connections reproduced low frequency information on its outputs better than the more spread network, while the more spread network reproduced high frequency information better (see Figure 3).

Despite this being the predicted result, it is a counterintuitive finding. Receptive fields with an on-off structure, or with full connectivity within the receptive field, do not show this pattern. I analyze these issues more in-depth in Appendix A of my thesis.

The model accounts for human reaction time differences, for left vs. right visual field presentation.

The main question of this research was whether these networks would encode images such that, when the model is given a classification task, the model shows similar left and right hemisphere biases as those
shown when humans perform that same task. For example, in the classic paper in this area (Sergent, 1982), subjects were asked to identify target letters when they were embedded in hierarchical letter stimuli (Navon stimuli) that were presented to the left or right visual field. Subjects showed faster response times for targets at the larger (global) level when the image appears in the left visual field (indicating a right hemisphere bias), while targets at the smaller (local) level elicit faster response times when presented to the right visual field (indicating a left hemisphere bias).

In order to test this, I trained the two networks on these stimuli, and recorded the hidden unit values, then I trained separate feed-forward classification networks on the classification task, using the hidden unit representations as inputs. In the case of Sergent’s study, the classification networks were trained to respond if there was a target at either the local or global level. The model hemisphere inputs led to different response biases in the classification network. Using a standard measurement of reaction time from such networks, the model showed the same visual field/hemispheric interaction as the subjects (see Figure 5).

This model was able to account for behavioral data in a similar way across a number of tasks, including hierarchical letter stimuli (Sergent, 1982), face recognition (Young & Bion, 1981), and a spatial frequency grating classification task (Kitterle, Hellige, & Christman, 1992). In each case, the network with the more narrow connectivity pattern performed like the LVF/RH, and the network with more widespread connectivity performed like the RVF/LH.

The model is developmentally plausible.

These results assumed the differential connectivity pattern is present in human visual cortex, but how might these arise developmentally? To answer this question, I created a model based on three developmentally plausible constraints: 1) The right hemisphere develops before the left (Geschwind & Galaburda, 1985; Hellige, 2006); 2) when the right hemisphere develops, the visual system is not mature, so inputs are blurry (see P. Wang and Cottrell (2013) for a review); and 3) during development, there is a great deal of synaptic pruning (Katz & Callaway, 1992).

In the model, left and right hemisphere networks began with identical connectivity—again, as many hidden units as pixels, and 12 connections per hidden unit. The networks were initially trained on blurry images with image clarity improving over learning, reflecting reflecting the development of the visual system. Intermittently during learning, the weakest connections were pruned, reflecting that long-range lateral connections experience both synaptic strengthening and connection pruning during development. Finally, the key manipulation was to begin the right hemisphere training with blurrier images than those initially used for the left hemisphere, reflecting a hypothesized difference in the timing of maturation between the hemispheres. Because it is hypothesized
that the right hemisphere begins maturing earlier, pruning would begin when visual acuity is poorer.

After training each network for the same amount of time, with slightly different bluriness of training images, I compared the connection distributions between the two models. Because the connection distributions began as identical, any difference in the distributions after learning could be attributed to a bias for learning blurrier (lower frequency) information. In order to visualize any difference between the distributions, I simply subtracted them (see Figure 6).

I found that the right hemisphere network, which was trained more on blurrier images, showed a bias for more narrow connections, while the left hemisphere network showed a bias for more widely spaced connections. This association validated the association found earlier between spatial frequencies and connection spread, when I had fixed the connection spread and allowed the network to learn the spatial frequency information.

Part II: Does timing affect interhemispheric integration?3

What causes the lateralization of function? One prevalent view is that lateralization can be driven by less communication between the hemispheres (Ringo et al., 1994; Rilling & Insel, 1999)—despite the fact that lateralized functions are often complementary (Gazzaniga, 2000; Hellige, 2006) and that interhemispheric functional networks are some of the strongest found (Stark et al., 2008). Recent experimental and modeling work on complementary lateralization of face and word processing (Plaut & Behrmann, 2011; Dundas, Plaut, & Behrmann, 2012) has suggested that interhemispheric interactions are critical in the development of lateralization in visual processing. In order to understand how this all works, I decided to carefully examine the two central papers suggesting that lateralization comes from decreased interhemispheric communication. This project examined and extended Ringo et al. (1994).

Long delays do not reduce interhemispheric communication.

Ringo et al. (1994), implemented a recurrent neural network with two model hemispheres interconnected by a model corpus callosum (see Figure 8). The model is trained to map random input patterns to random output patterns, so that the output pattern is on after T time steps. All connections in the model had a time delay of 1 except the

3Presented at COGSCI 2013, Society for Neuroscience 2014, and SRCD 2015
Figure 9: (left) Ringo et al. (1994)’s original model data. (right) The same data, with the Delay=10 curve shifted to the right by 9 time steps. This shift compensates for the 9 time step difference in onset to communication. The overlap of the right figure indicates no decrease in communication beyond the onset difference.

interhemispheric connections. In one condition these connections had a time delay of 1; in the other, they had a time delay of 10. After training, interhemispheric dependence was measured by removing the interhemispheric connections and computing how much more error the network produces without the connections vs. with them. This is called “lesion-induced error” (Lewis & Elman, 2008).

Ringo et al. found that when the total run time of the network (T) was less than 35 time steps, the network with longer delays showed less lesion-induced error. They interpreted this to mean that interhemispheric communication is selectively reduced by long time delays.

My first analysis was simple, and is depicted in Figure 9. We can expect a difference of 9 time-steps between the networks for interhemispheric communication to begin. After accounting for this obvious difference, what other differences remain? I shifted the “Delay 10” curve by 9 time steps to match the onsets between the curves, and I found that the curves completely overlapped. In other words, there is no difference in interhemispheric communication between the two delays, except for a delay in its onset. Once that onset has passed, the two networks seem to communicate identically over the interhemispheric connections.

There are a number of reasons to think that the onset delay has a much lesser effect in real brains doing everyday things than the small effect demonstrated in the paper. First, the time difference between intra- and interhemispheric communication in the model is 10-fold–much greater than what would be expected in a real brain (estimated to be about 5 - 25 milliseconds). This is also mitigated because real brains have a wide distribution of axon diameters; thus some information arrives extremely quickly, regardless of brain size (Olivares, Montiel, & Aboitiz, 2001). Second, the brain can overcome delays through temporal correlations and prediction. A signal arriving over a delay may still hold a lot of relevant information (be temporally correlated). In addition, a hemisphere can learn to predict the upcoming input (also through temporal correlations), and
therefore send predictive information across the delay to compensate. In these simulations, because the model network cannot predict when an input will be shown, nor what it will be, the onset delay is seen. In most cases of cognition in the wild, there is high predictability (high temporal correlation) within the very short time delay difference (5 - 25 milliseconds) for most tasks. Thus, any claim that conduction delay affects interhemispheric communication should be very careful to specify why there are no temporal correlations in that particular context that would overcome it.

**Conduction delay variability reduces interhemispheric communication.**

The network overcomes delays by using temporal prediction, allowing one hemisphere to utilize the delayed activity of the other when it arrives. Therefore, any decrease in temporal correlations as information is transmitted should lead to decreased reliance on the slow interhemispheric connections.

One potential source of noise is from long, thin, unmyelinated axons. These axons have been shown to have unreliable conduction delays through recordings (S. S. H. Wang, 2008) and through detailed biophysical simulations (Faisal, Selen, & Wolpert, 2008). The degree of variability is a function of delay magnitude—longer axons cause both longer delays and more transmission noise. Electron microscopy studies in monkeys (LaMantia & Rakic, 1990a) and cats (Berbel & Innocenti, 1988) indicate that, early in development, the preponderance of callosal fibers are long, thin, and unmyelinated. These fibers increase their diameter and become myelinated over development (see Figure 11).

To test the effect of delay unreliability, I ran simulations using a more powerful version of of Ringo et al’s model (Lewis and Elman (2008); see Figure 8), but introduced delay-dependent noise on all of the connections\(^4\). The noise was a linear function of the connection’s activity (only active neurons are noisy) and the connection’s delay magnitude (longer delays are noisier).

I found that this delay-dependent noise decreased lesion-induced error for the network with long interhemispheric delays, and had little effect on the network with short interhemispheric delays (see Figure 10). Even

\(\text{noise} = 1\%\)

\(\text{(training error)}\)

\[^4\text{The data and analysis code are available at http://github.com/guruucsd/NoisyCC.}\]
after shifting the results to account for an onset delay, a gap remained in lesion-induced error, indicating that the two connected networks learned to ignore the unreliable input from the other hemisphere. Hence the noise led to a reduction in interhemispheric communication selective to long delays, such as would be expected in the thin, unmyelinated axons found in the corpus callosum at birth.

Neural maturation reduces conduction delay variability, changing functional circuits over development.

Over development, functional networks gradually reduce their bias for local connections and incorporate long-distance connections (Kelly et al., 2009; Uddin, Supekar, & Menon, 2010). To investigate if this model would show a similar effect, I simulated maturation by slowly reducing the delay-dependent noise during training and examined how lesion-induced error changed over training. When noise was fully present, lesion-induced error increased as compared to a network trained with no noise (see Figure 12). As noise decreased, the lesion-induced error trended more and more to that of the no-noise network. When noise was completely turned off, the lesion-induced error in the previously noisy network slowly converged to that of the no-noise network, suggesting that the two model hemispheres were more interdependent. This mirrors the findings that functional networks increasingly include long-range connections over development.

Part III: How does connectivity change with brain size?5

In this project, I addressed the second key paper (Rilling & Insel, 1999) claiming that lateralization in humans is caused by larger brains having more independent hemispheres. In that paper, Rilling and Insel claim to show proportionally decreased anatomical connectivity over the corpus callosum with increasing brain size, which they suggest also leads to functional lateralization.

---

5Presented at the Bernstein Conference 2013 and Society for Neuroscience 2014
Are surface areas a good proxy for connection counts?

Rilling and Insel (1999) suggested that examining the ratio between connectivity within and across the hemispheres is a good way to estimate interhemispheric functional connectivity across species. In order to understand whether this ratio is reduced in large-brained species such as humans, they collected full brain MRI volumes across 11 species of primates to estimate grey matter surface area (GMSA), their proxy for total neurons and therefore total connectivity, and callosal cross-sectional area (CCA), their proxy for interhemispheric connectivity. They compared these surface area data using allometric regression, which is simply linear regression on log-transformed data, appropriate for the scale-free power-law relationships that tend to best fit these kind of cross-species scaling relationships.

Rilling and Insel found that the CCA increases sub-linearly with GMSA (exponent: 0.88). This suggests that as brains gets bigger, the pace at which total white matter grows is increasingly faster than the pace at which interhemispheric connectivity grows. They concluded, based on this analysis, that interhemispheric connectivity is selectively reduced in larger brained species, leading to decreased interhemispheric communication and therefore greater asymmetry.

I aimed to estimate the number of white matter connections and number of callosal connections across species directly. The simple equations for estimating fiber counts from Rilling and Insel’s data are found below. Because each of these quantities have their own regression exponents, Rilling and Insel’s surface-based estimates make strong, unlikely assumptions about the other quantities here. In fact, all of the quantities needed to estimate the connection counts have allometric regressions published in the literature, except one—and that one (callosal axon density) has raw data published that could be used to compute the regression.

Formulas:

\[
\text{[# callosal connections]} = \text{[callosal area]} \times \text{[axon density / mm}^2]\]
\[
\text{[# white matter connections]} = \text{[# cortical neurons]} \times \text{[% projecting]}\]
\[
\text{[# cortical neurons]} = \text{[grey matter volume]} \times \text{[neuron density / mm}^3]\]

Note: bolded values had previous allometric regressions published.

In order to access all of the data used in these analyses, I used some basic computational techniques to parse out raw data from published figures. In all, I created a database of over 15 publications containing the cross species data used in this analysis—all data that were otherwise inaccessible—including the data from Rilling and Insel6.

6The data and analysis code are available at http://github.com/guruucsd/CallosalScaling'.
Callosal fiber density decreases with brain size.

To compute callosal fiber density, I used data from six small-brained species (S. S. H. Wang et al., 2008) to compute an allometric regression of callosal fiber density and brain weight. I found a negative power-law association between brain mass and fiber density (exponent: -0.28), consistent with scattered, lower-quality samples in the literature.

I wished to use this result in estimating callosal fiber counts for all species, but that would require extrapolation well beyond the data used in the regression. In order to validate that this regression is a good estimate for large-brained species, I found human data to compare with the predictions of this curve. However, the best human data available (Aboitiz et al., 1992) had a number of confounds, including a large age difference (both in absolute and lifespan) compared to Wang’s et al.’s animal data. The most relevant available data (monkeys (LaMantia & Rakic, 1990b) and humans (Aboitiz, 1991)) showed a clear age-dependence of callosal fiber density.

Human data fit well with animal data; this allows missing human data to be estimated.

Two steps were necessary to correct for the age differences in the samples. First, I estimated an aging curve for callosal fiber density, using high-resolution data from macaques (LaMantia & Rakic, 1990b). Second, I used a “landmark” technique for mapping lifespan ages from humans to macaques and back again (Finlay, Darlington, & Nicastro, 2001). I determined a correction factor of 1.2 to map the human data from the average age of 45 to an earlier, “young adult” age.

After applying all corrections, the human data fit quite well with the prediction made by the allometric regression on small-brained species (Figure 14). Using this age correction on the human data, I conservatively estimate that young adult humans—a more standard age to report—have on average 240 million fibers in their corpus callosum.

A similar technique could be used to estimate the number of fibers in the human corpus callosum at birth, using more data from LaMantia and Rakic (1990a). My best estimate is 800 million fibers—a number that has been grossly underestimated by data collection techniques using subpar optical imaging (140 million, reported by Luttenberg (1965)).
The surface-based estimates actually underestimated the cross-species scaling effect.

After successfully estimating regressions for all of the above quantities and plugging them into the above equations, I found that callosal connectivity scales at a much slower rate of total connectivity than Rilling and Insel reported (exponent: 0.64, vs. 0.88 estimated by Rilling and Insel, 1999). This difference in scaling is dramatic; the previous estimate is nearly linear, but the current estimate is drastically sublinear. Given that we know that human interhemispheric communication is extremely robust, it seems unlikely that comparing total intrahemispheric and interhemispheric connectivity is a good way to estimate how interhemispheric functional connectivity changes across species.

There is no cross-species scaling difference for a more functionally relevant measure.

To find an explanation, I focused on the scaling of fiber counts in inter-area fiber bundles. Cortical areas vary in their interconnectivity, and the strength of these fibers is often an indicator of the functional relationship between them (Markov et al., 2013). Two inter-area connections within a hemisphere presumably have different functions, and so it is somewhat meaningless to aggregate them. Instead, I proposed that examining the scaling of the number of fibers per intrahemispheric vs. interhemispheric inter-area fiber bundles would be a better indication of how functional networks change with brain size. If interhemispheric connectivity is selectively reduced with brain size, as Rilling and Insel suggest, then left V1 → right V1 should be proportionally smaller and smaller compared to left V1 → left V2 or left V1 → left MT.

With this comparison of inter-area fiber bundles in mind, differences between intrahemispheric and interhemispheric connections immediately are apparent. A given cortical area will connect with a large number of other cortical areas in the same hemisphere. In addition, as brains get bigger, the total number of cortical areas—and the number of cortical areas that a single area connects to—increases (Changizi & Shimojo, 2005). The story is completely different for interhemispheric inter-area connections. In small-brained mammals and large-brained mammals alike, one cortical area will connect almost exclusively with only one other cortical area in the other hemisphere—the homologous area.

In order to see if this difference fully explains the aggregate results above, I computed two ratios and compared them using allometric regression. The first ratio is the proportion of callosal fibers vs. intrahemispheric fibers, as computed previously. The second ratio is the number of interhemispheric inter-area fiber bundles vs. the number of intrahemispheric inter-area fiber bundles. If these two ratios scale identically, then the differences in fiber counts can be completely attributed to the differences in the number of inter-area connections, as brains get bigger.

Indeed, the allometric regression between the two ratios produced an exponent of 1.04, indistinguishable from
unity. This indicates that the number of fibers within an inter-area fiber bundle scales similarly, regardless of whether the cortical areas connected are within the same hemisphere or across hemispheres. Using the same set of data, I could easily estimate the number of fibers per intrahemispheric and interhemispheric inter-area fiber bundle. I estimated 3 to 8 times more fibers in an interhemispheric fiber bundle as compared to an intrahemispheric fiber bundle. This result is even more stunning when we consider that the number of connections in a fiber bundle generally exponentially decays with distance, and the average interhemispheric inter-area connection is longer than the average intrahemispheric inter-area connection. This suggests that interconnectivity across the corpus callosum may have a very special role compared to the numerous intrahemispheric connections that a cortical area makes.

Conclusions

In my dissertation research, I used neural data to construct computational models about issues relevant to cognitive science–vision, lateralization, and interhemispheric communication. Because I focused on specific neural connections, I could use knowledge specific to those connections in order to extend my models to development. I credit this interdisciplinary interplay between neuroscience, modeling, and behavior for the number of specific and testable hypotheses about the microstructure of human cortex that were generated by my research. These predictions can be further refined through cost-effective computational modeling. When costs and collaboration allow it, these hypotheses can be tested through more expensive post-mortem tissue analysis.

These predictions include:

- An asymmetry in the spatial spread of long-range lateral connections, likely in lateral-occipital cortex.
- The fiber count in the corpus callosum of human infants (800 million fibers) and young adults (240 million fibers).
- Relative fiber counts for inter- and intra-hemispheric fiber bundles connecting cortical areas (3 to 8 times greater for interhemispheric bundles).

References


