Enterprise Biology Software: XIII. Research (2012)

ROBERT P. BOLENDER

Enterprise Biology Software Project, P. O. Box 292, Medina, WA 98039-0303, USA
http://enterprisebiology.com

SUMMARY

The topic of the report this year is "Complexity theory applied to clinical diagnosis of the brain." In short, we will work through the steps of extending the clinical application of MRI from largely a descriptive role (radiologist views images and writes reports) to an analytical one (software uses MRI data directly to diagnose disease). The first part of the report describes a database that treats the brain as a complexity. It captures disorders as mathematical markers by forming numerical ratios (X:Y:Z) accompanied by three named parts (AX:BY:CZ). The raw data used to calculate these markers came from an MRI database (Internet Brain Volume Database (IBVD)) developed as part of the Human Brain Project (Kennedy et al., 2012). The process consists of taking 1,630 data points from this database and then satisfying the large numbers requirement of complexity theory by transforming them into more than 700,000 mathematical markers. This gives us the phenotypic properties of twenty-four brain disorders, each expressed as a complexity in terms of the parts and connections. Our first task will be to review the protocols developed for using these markers to diagnose disease. In short, they consist of identifying an unknown disease by entering published patient data into the diagnosis database and then matching the unknown markers to known standards. In turn, the results are tallied and the diagnosis goes to the disorder with the highest score. The results thus far are most encouraging in that the software consistently delivers the same diagnosis as the physicians.

The second part of the report consists of analyzing the mathematical markers graphically in an attempt to uncover fundamental patterns associated with disease. Here our goal becomes the development of a theory structure for disease based on complexity. The analysis reveals that the brain uses a modular approach when assembling both normal and abnormal brains. Data pairs, triplets, and now mathematical markers all detect a well-defined biological order (stoichiometry) based on parts and connections (Bolender, 2001-2012). A new and potentially disruptive finding is that different diseases in the brain often display striking similarities. Graphics clearly show that a considerable overlap exists in the parts and connections of many diseases, thereby suggesting common etiologies. Consider, for example, schizophrenia. Given its vast array of markers, it may be the most complex disease of the central nervous system – or at least the one with the largest published data set. When its complexity is unfolded graphically, schizophrenia contains exact replicas – qualitatively - of at least six other diseases and close replicas of several more. Is there a quantitative explanation? Apparently, yes. Schizophrenia shares large numbers of identical mathematical markers with those of its "embedded" diseases. The larger picture to emerge from this analysis is that the disease process in the brain occurs as a well-ordered event, wherein putative mistakes in design combine to produce new diseases with new emergent properties. Perhaps the most unexpected finding of all, however, comes from the original MRI database (IBVD).

Since each mathematical marker consists of six variables, it seems remarkable indeed to discover that different papers routinely generate exactly the same mathematical markers (the current record stands at 36 duplicates). Given this new found ability to phenotype the human brain in health and disease, we now have the option of using MRI data directly to diagnose diseases of the brain. Considering the enormous complexity divide that currently exists between genes and phenotypes, patient care based on phenotypic data carries the potential of creating a new information industry from patient data, one that can contribute importantly to the emergence of an evidence-based medicine.

INTRODUCTION

Exploring the unknown in science requires a theory structure capable of defining the guiding principles, rules, and procedures that together create an environment consistent with new discovery. In biology, the experimental method and reductionism have advanced the science to its current level with resounding success. Together they have allowed us to accumulate vast amounts of new information by taking biology apart and characterizing its parts in great detail - up and down its hierarchy of size. As a result, the biology enterprise has been built largely around isolated data points wherein change is usually detected by comparing one point to another or by fitting points to regressions. This isolation, however, creates problems.

Recall that the guiding principle of reductionism is simplification, wherein a very small sample is taken from a much larger whole. In exchange for such access, we have been willing to forfeit the complexity of the whole. Unfortunately, many of the problems we are trying to solve today require the complexity we no longer have. One way of dealing with this shortcoming is to recover the lost complexity and then use it to address our most challenging problems. The report explains how this can be done.

We begin by considering a new theory structure. Complexity theory, which focuses on the whole, comes with its own set rules, data, and procedures (Bolender, 2011). Like any new theory in science, the rules come to us - one by one - from the data as we use them to test our new ideas and findings. To get started, however, we have to satisfy a few preliminary requirements. First, we need to identify the biology literature as a primary data source and then generalize the literature across papers, data types, and settings. We can do this by defining the elemental unit of complexity as a dimensionless ratio, one that carries information about the parts and their connections (Bolender, 2001-2011). The final preparative step involves change. In complexity, change is detected by identifying alterations in patterns, not by just comparing the value of one part to another. Moreover, change behaves as a complexity that ripples throughout the biological hierarchy, often involving staggering numbers of parts and connections. This wide-ranging connectivity provokes the overarching rule of complexity theory, which states that it takes a complexity to solve a complexity. How does this rule help us? If, for example, we use it to create a quantitative complexity, then we can reduce a biological problem to a mathematical problem. This is exactly what the report does. It treats diagnosis and disease as complexities and then proceeds to solve them.

Although the task of applying complexity theory to biology may sound somewhat daunting, it is surprisingly easy to do and understand. To many it will seem a logical next step toward understanding those things that currently exist beyond the reach of reductionism, our current theory structure.

Consider a simple example of how complexity theory works. Going from reductionism to complexity is akin to going from chaos to order. We can see this transformation occur by plotting data taken from the Internet Brain Volume Database (IBVD). Figure 1 illustrates 61 estimates (control and experimental) for the volume of the amygdala, sorted according to size. The chaos appears in the figure as a broad range in the individual values plus the expected separation produced by the data carrying different units (mm^3 and cm^3).

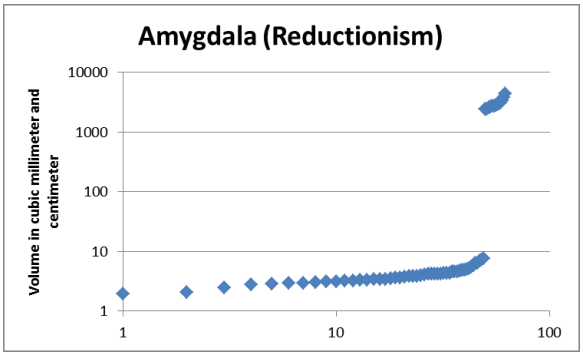

Figure 1. MRI estimates of volume for the human amygdala. When biological data become isolated, they no longer have their connections and become chaotic.

If we add back the connections, by creating ratios between the parts of the amygdala, the original order begins to reappear (Figure 2).

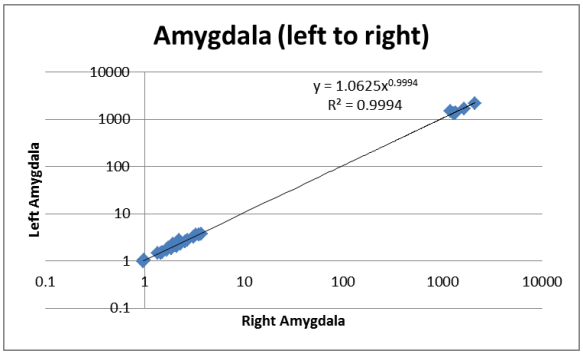

Figure 2. Now the MRI data of Figure 1 - expressed as ratios - display order, as detected by a power regression with an R^2=0.9994.

Next, we calculate a triplet ratio (X:Y:Z) wherein X is set equal to one. This allows us to express the Y, Z data as a repertoire equation (Figure 3).

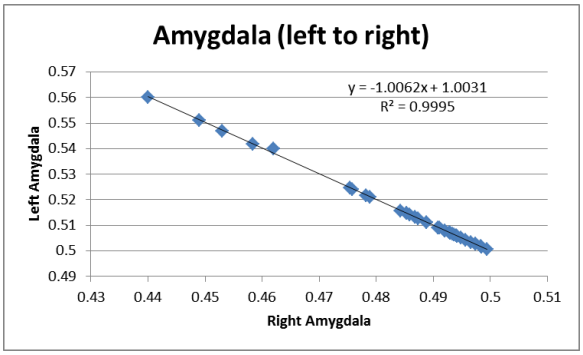

Figure 3. After forming the ratio X:Y:Z where X=1, Y (left) is plotted against Z (right). The result is a linear equation with an R^2=0.9995.

Notice in Figure 3 that the range of the data has been compressed by orders of magnitude (compared to Figure 1) and that the two parts observe a strict linearity. Now, most of the data differ by less than five percentage points.

If Figure 1 represents chaos and Figure 3 order, then we want to be somewhere in between near the edge of chaos, the place where the most interesting things happen. This final step consists of going from repertoire values (Figure 3) to decimal repertoire values (Figure 4). A remarkable thing happens. All the data shown in Figure 1 condense into a single ratio (0.4:0.5). The point of this example is to show how complexity theory can optimize the effectiveness of our data. Notice that reductionist data are noisy and difficult to explain (Figure 1), whereas the same data in a complexity setting (Figure 4) become wonderfully quiet and explain themselves as a generalization reflecting the underlying design principle of biology.

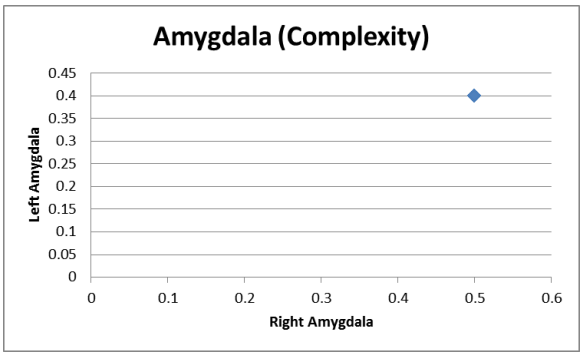

Figure 4. A single decimal repertoire value (ratio) for the amygdala (Y:Z) - left vs. right – summarizes the amygdala data across many different publications and diseases. Notice that all 61 data points are now represented by a single ratio (Y:Z = 0.4:0.5). The mathematical marker - *amygdala1amygdalaleft0.4amygdalaright0.5* – can be found in the MRI database.

METHODS AND RESULTS

The software package for 2012/2013 includes new software tools for assembling, analyzing, and interpreting mathematical markers based on MRI volume data (Keller and Roberts, 2009). Special attention will be given to diagnosing disease within the framework of a preliminary theory structure based on complexi-

ty. The challenge for the reader will be to learn how to explore a complexity by creating, managing, and interpreting large data sets. Patterns rather than individual data points become the focus of complexity wherein connectivity translates local changes to global consequences.

Enterprise Biology Software Package

The software package includes eight screens offering ready access to programs, databases, and documents (Figure 5).

Figure 5. Enterprise Biology Software Package – 2012/13.

Generating Mathematical Markers

Mathematical markers consist of a six character string that includes three named parts (A,B,C) and three numerical values (X:Y:Z) expressed as a ratio (AX:BY:CZ). The method for generating these markers can be summarized briefly. Beginning with a database of published MRI data (Internet Brain Volume Database (IBVD)) of the Human Brain Project (Kennedy et al., 2012), the data set of a given paper is expanded by forming permutations to produce two sets of triplets: A,B,C and X:Y:Z. In turn, the numerical ratio is converted into a decimal ratio and the alpha numeric components (A,B,C and X:Y:Z) are concatenated to form a mathematical marker (AX:BY:CZ). Using this procedure, fewer than 2000 individual data points were transformed into more than 700,000 markers.

Next, we consider a step by step example. Figure 6 illustrates a collection of named parts (a, b, c) entered into the permutations function of Mathematica (Wolfram Research, Inc.). Figure 7 deals with the numerical values.

In[3]:=
```
Permutations[{a, b, c}, {3}]
```
Out[3]=

$$\begin{pmatrix} a & b & c \\ a & c & b \\ b & a & c \\ b & c & a \\ c & a & b \\ c & b & a \end{pmatrix}$$

Figure 6. Generating all possible combinations of three named parts, a, b, and c - taken three at a time. In practice, the letters are replaced with the names of brain parts and corresponding ratios are concatenated thereto.

In turn, the three columns of named parts are copied from Mathematica and pasted into the columns (X Name, Y Name, Z Name) of a spreadsheet template (Figure 8); C:\EBS 2012\Files\TEMPLATE.

The procedure is repeated for the numerical values (Figure 7). The resulting three columns of numbers are copied from Mathematica and pasted into the three value columns of the template spreadsheet (X Value, Y Value, Z Value). As soon as the data are entered, the ratio columns (X Ratio, Y Ratio, Z Ratio) are calculated automatically by the spreadsheet (Figure 8).

In[1]:=
```
Permutations[{1, 2, 3}, {3}]
```
Out[1]=

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \\ 2 & 1 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \\ 3 & 2 & 1 \end{pmatrix}$$

Figure 7. All possible combinations of three numerical values (1, 2, 3) taken three at a time.

The final step of the procedure consists of expressing the ratio data as decimal ratios – using the numbering scheme identified in the blueprint software and supplied as a form: Documents > Forms > Worksheet – Connection Phenotype (Bolender, 2007-2012). Since the values in the X Ratio column are set equal to one by convention, values are needed for just the Y and Z decimal ratio columns. First, the Y Ratio column is sorted low to high and then the Y Decimal column data are entered (according to the numbering scheme) and the completed column is copied to a worksheet adjacent to the main data table (Figure 8). Next, the Z Ratio column of the template worksheet is sorted low to high and the Y column stored in the adjacent worksheet is copied and pasted into the Z Decimal Ratio column. The mathematical markers appear automatically and data entry is complete, as shown in Figure 8. Using the spreadsheet template and the procedure described above, thousands of markers can be generated in just minutes. See EBS 2012/Setup/Read.

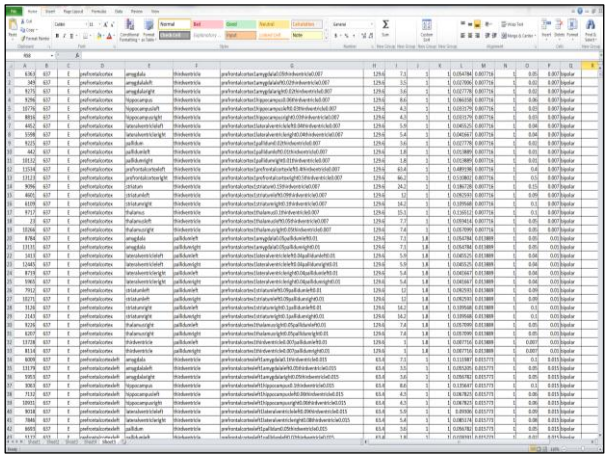

Figure 8. The spreadsheet template used to generate mathematical markers. Enlarge as needed.

Mathematical Marker Data Sets

Starting with the IBVD database of MRI volumes (Kennedy et al., 2012), the data of 67 publications were converted into mathematical markers (normal and disease) and stored as spreadsheets and relational database tables (normal, disease, normal+disease, and diagnosis); see Table 1.

Table 1. Four MRI data sets (mathematical markers) are available as spreadsheets and relational database tables.

Name	Spreadsheet	Database Table	Markers
Normal	●	●	295,404
Disease	●	●	428,784
Normal + Disease	●	●	724,188
Diagnosis	●	●	61,204

The first three data sets listed in Table 1 (normal, disease, normal+disease) group markers collected paper by paper. The diagnosis data set contains the disease data minus the disease data that have duplicates in the controls. These false positives account for 367,580 markers, reducing the number of diagnostic markers to 61,204.

When dealing with large and complex data sets, we clearly increase our options by moving these data back and forth between spreadsheets and the tables of a relational database. This is accomplished as follows. To move the markers from a spreadsheet into a relational database, save the spreadsheet as a tab delimited text file (e.g., diagnosis.txt), copy it to the root directory of the target computer (e.g., C:/diagnosis.txt), open the MRI view screen of the diagnosis database (e.g., Figure 9), and click on the **import** button. To add data from a spreadsheet to the diagnosis database, number the new rows of the spreadsheet consecutively (beginning with 61,205), save it as a tab delimited text file in the root directory (C:/diagnosis.txt), and import it by clicking on the **import** button. To save, click on the **update** button. To convert a database table into a spreadsheet, click on the **save as** button (Figure 9), name the file, and save it as an Excel file (*.xls). Alternatively, a database table can be saved as a text file (*.txt), which can be copied and pasted into an Excel worksheet.

For these data transfers between spreadsheets and databases to work smoothly, we need several options for deleting rows of data from the database tables (Figure 9). They include deleting all rows, all rows associated with a given citation, all rows having numbers greater than 61,204, and one row at a time. Command buttons attached to the view data screen of the diagnosis table (Figure 9) run these operations automatically.

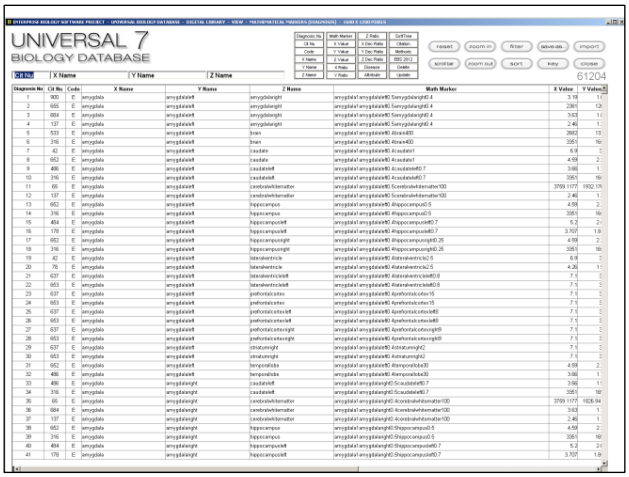

Figure 9. Mathematical markers of the diagnosis data set are shown in the view data screen of the MRI database.

For convenience, a query by example (QBE) frontend is included with the database tables (Figure 10). It allows us to run complex queries and to find specific information quickly – even with large data sets.

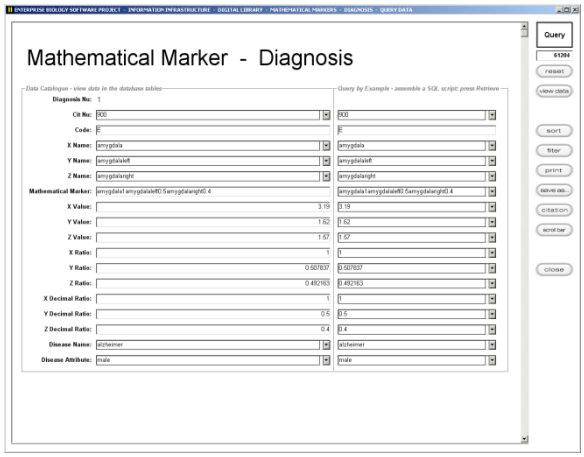

Figure 10. The screen represents a QBE frontend to the diagnosis database. It provides a user-friendly way to assemble complex SQL scripts automatically. The results can be viewed directly as a scrolling table or sent to an Excel spreadsheet.

Diagnosis Database

A mathematical marker represents a basic unit of complexity, which in sufficiently large numbers allows us to capture the complexity of the human phenotype in remarkable detail. To illustrate this point, we will use the markers to assemble and test a software-based approach to clinical diagnosis. The key player is the diagnosis data set, which contains known markers for 24 different disorders of the brain. The diagnostic method consists of generating mathematical markers for an unknown disorder, adding them to the diagnosis data set, marking exact matches (unknown marker = known marker), tallying the results, and making the diagnosis. Although the diagnoses were originally done with spreadsheets, the results can be transferred to and viewed more conveniently with database tables (Figures 11 and 12).

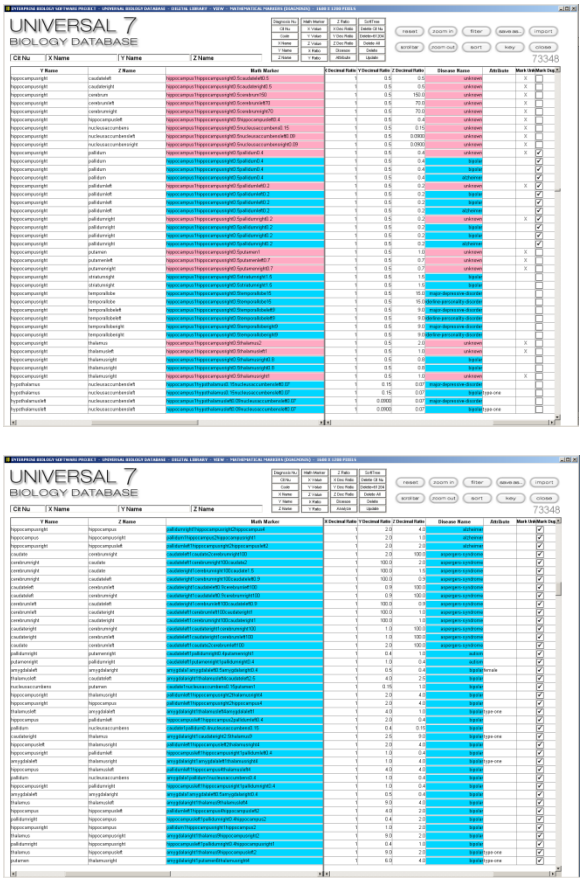

Figure 11. Diagnosing an unknown disorder. Top: The diagnostic procedure consists of mixing the markers of 24 known disorders (blue) with markers coming from an unknown disorder (red) and checking off duplicates that occur between known and unknown markers. Bottom: When all the duplicates have been identified, clicking on the analysis button gives the results. The disorder with the largest number of hits becomes the diagnosis.

At this point, it may be helpful to work through the main steps of generating the standards for the diagnostic procedure. The process begins by select-

ing a publication, identifying the data, and then using the data to generate collections of mathematical markers. This provides two spreadsheet tables, one for normal patients (C) and the other for those with a diagnosed disorder (E). Since the diagnosed patients carry both normal and abnormal markers, we need to isolate the collection of markers unique to disease. This is accomplished by combining both types of markers (C+E) in the same Excel spreadsheet, sorting on the markers column, and searching the list for duplicates. Figure 10 illustrates the result of such a search.

right15caudate10	631	11378	8404	1	18.0317	13.3185	1	15	10	major-depressive-disorder
right15caudateleft6	705	11902	4305	1	16.8823	6.10638	1	15	6	
right15caudateleft6	631	11378	4134	1	18.0317	6.55151	1	15	6	major-depressive-disorder
right15caudateright6	705	11902	4351	1	16.8823	6.17163	1	15	6	
right15caudateright6	631	11378	4270	1	18.0317	6.76704	1	15	6	major-depressive-disorder
right15hippocampus10	705	11902	7115	1	16.8823	10.0922	1	15	10	
right15hippocampus10	631	11378	6958	1	18.0317	11.0269	1	15	10	major-depressive-disorder
right15hippocampusright5	705	11902	3680	1	16.8823	5.21986	1	15	5	
right15hippocampusright5	631	11378	3581	1	18.0317	5.67512	1	15	5	major-depressive-disorder
right15hypothalamus2	705	11902	1410	1	16.8823	2	1	15	2	
right15hypothalamus2	631	11378	1275	1	18.0317	2.0206	1	15	2	major-depressive-disorder
right15hypothalamusleft1	705	11902	705	1	16.8823	1	1	15	1	
right15hypothalamusleft1	631	11378	644	1	18.0317	1.0206	1	15	1	major-depressive-disorder
right15nucleusaccumbens1.5	705	11902	1240	1	16.8823	1.75887	1	15	1.5	
right15nucleusaccumbens1.5	631	11378	1086	1	18.0317	1.72108	1	15	1.5	major-depressive-disorder
right15nucleusaccumbensleft0.8	705	11902	615	1	16.8823	0.87234	1	15	0.8	
right15nucleusaccumbensleft0.8	631	11378	535	1	18.0317	0.84786	1	15	0.8	major-depressive-disorder
right15nucleusaccumbensright0.8	705	11902	625	1	16.8823	0.88652	1	15	0.8	
right15nucleusaccumbensright0.8	631	11378	551	1	18.0317	0.87322	1	15	0.8	major-depressive-disorder
right15putamenleft7	705	11902	5263	1	16.8823	7.46525	1	15	7	

Figure 10. The spreadsheet table identifies duplicates between control (green) and experimental (white) rows. Such rows are deleted to avoid false positives when making a diagnosis.

By deleting all the duplicates, the remaining markers of a given paper are unique to the disease. In effect, this operation removes the false positives at the local level, one paper at a time. However, when all the markers from many papers are combined in a single spreadsheet (the diagnosis data set), an abnormal marker in the data set of one paper may have a normal counterpart in another. By combining – in a single spreadsheet - the entire normal data set with the all the locally filtered abnormal markers, duplicates appearing at the global level can be identified and removed.

This procedure produced the diagnosis database, which contains 61,204 mathematical markers unique to the 24 diseases contained therein. Next, we can test the effectiveness of the database as a diagnostic tool by working through several representative examples.

Example 1: Unknown Disease

The first test consists of generating a collection of unknown mathematical markers using the data set of a published paper not found in the diagnosis database. These unknown markers are highlighted in yellow, added to the diagnosis spreadsheet, and matched to known markers by identifying duplicates (Figure 11). All 61 matches identifed the unknown disorder correctly as schizophrenia.

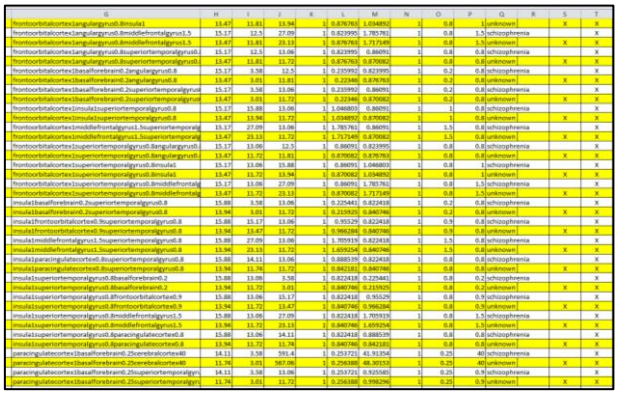

Figure 11. The software-based diagnosis consists of matching known to unknown markers. The results indicate that all 61 matches identify schizophrenia correctly as the unknown disorder.

Example 2: Unknown Disease

Using a different paper as a data source, the procedure described in Example 1 is repeated. This time, however, the diagnosis identified seven different candidate diseases with bipolar receiving the largest number of matches (Figure 12). Although reassuring, because bipolar is the correct diagnosis, the results clearly show that the same mathematical marker can appear in different diseases. Indeed, such an observation is more often the rule than the exception. This alerts us to the possibility that the sample used for a diagnosis may influence the outcome. A relatively small number of parts, for example, may not provide a sufficiently large number of unique markers to make the diagnosis correctly. Thus far, tests of unknown diseases represented by moderate to large data sets (>=15 parts) were not found to be subject to this limitation. This is likely to become less of an issue as we begin to identify the optimal composition of data sets for diagnosis.

1	6	0.4	unknown		x	
1	6	0.4	unknown		x	
1	6	2	unknown		x	
1	6	1	unknown		x	
1	6	1	unknown		x	
1	6	3	alzheimer	female		x
1	6	3	bipolar	type-one		x
1	6	3	unknown		x	x
1	6	3	alzheimer	female		x
1	6	3	bipolar	type-one		x
1	6	3	unknown		x	x
1	6	9	unknown		x	
1	6	4	alzheimer	female		x
1	6	4	bipolar	type-one		x
1	6	4	unknown		x	x
1	6	4	alzheimer	female		x
1	6	4	bipolar	type-one		x
1	6	4	unknown		x	x

Figure 12. Notice that one marker can apply to more than one disease. For example, alzheimer, bipolar, and unknown all share the same mathematical marker (amygdalaright1putamen6putamenleft3) having the ratio 1:6:3.

Figure 13 summaries the results of Example 2.

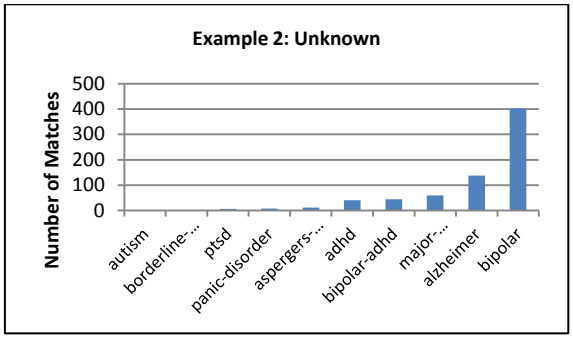

Figure 13. The figure identifies the disorders that share mathematical markers with the unknown disorder, which in this case is identified as bipolar. The analysis indicated that bipolar disorder had 712 markers of which 401 (56.3%) were uniquely bipolar.

Figure 14 summarizes the data of Figure 13 with an equation, which can define a disease as a unique quantitative pattern.

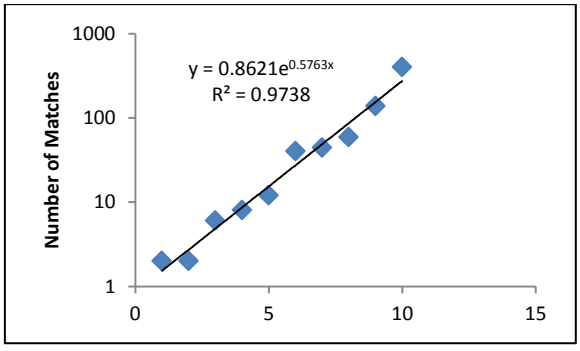

Figure 14. Disorders sharing mathematical markers with bipolar disease are plotted as an exponential. In effect, this is the equation for the bipolar disorder based on mathematical markers and the current data set.

Example 3: Alzheimer Disease

Next, if we look at the mathematical markers of Alzheimer disease in the diagnosis data set, we see a pattern remarkably similar to the one shown in Figure 13. The disease shares 878 of its 4392 markers with 12 other diseases, ranging from bipolar (606) to epilepsy (1) – see Figure 15. This pattern of sharing mathematical markers with other diseases appears as a general feature of brain disorders and may help to explain the difficulty often encountered when making a diagnosis based largely on symptoms.

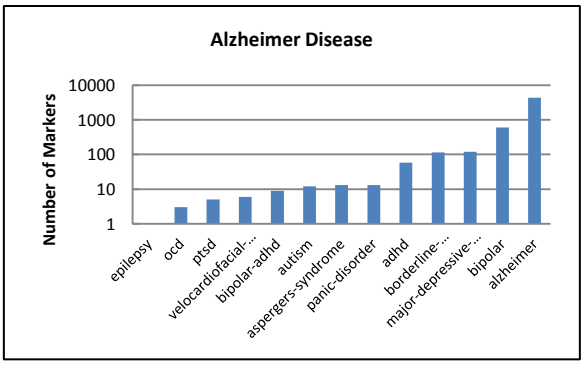

Figure 15. Several diseases share the same mathematical markers with Alzheimer disease. Of the 4392 markers for Alzheimer disease, 3514 (82%) are unique, whereas 606 (11%) are shared with bipolar disorder, 121 (2%) with major depressive disorder, 115 (2%) with borderline personality disorder, and 58 (1%) with ADHD. This pattern of sharing mathematical markers is typical of brain disorders.

Thus far, the examples demonstrate that complex disorders unfold into well-defined collections of basic building blocks (or modules) that can be detected as mathematical markers. This unfolding process is important because it provides insights into the rules of biological complexity. If, for example, we reverse the process and begin with the modules, then we can readily assemble disorders that mimic those of biology. We can do this because the unfolding process reveals the recipe of a disorder as a specific construct of parts and connections (Figures 13-15). The principal insight into disease as a complexity emerges as a general rule of design. Parts and connections are arranged into modules (e.g., markers), which, in turn, are arranged into larger patterns that define a disorder. This modular ap-

proach is an ingenious one on the part of the brain in that it can use many of the same parts and modules to create a wide range of different disorders. Moreover, it allows us to generalize disease as a global collection of markers, with local subsets expressing specific disorders. The diagnosis database performs both of these functions simultaneously. Let's look at an example.

By unfolding the diagnosis database, we can see the relationships of parts to disorders (Figure 16). Notice that a disorder is defined by its parts, that different disorders often share the same parts, and that a relatively small number of parts (35/185 or 19%) accounts for much of the damage that occurs in the brain.

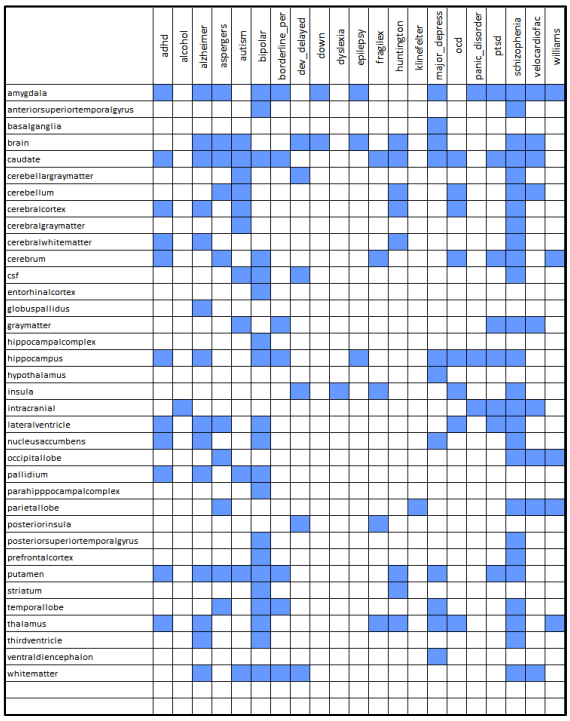

Figure 16. Summary of brain parts involved in the disease process. The data come from the diagnosis database. Enlarge as needed.

Example 4: Autism

Visualizing change in a complexity creates a challenge for the following reason. When parts are connected, a local change in one or a few parts typically spreads throughout the complexity. In effect, we need ways of observing changes that occur in very large numbers of parts – often numbering in the tens of thousands to millions.

Recall that in calculating data ratios (X:Y:Z), X is set equal to 1.0. This simplifies our task of detecting patterns of change somewhat by only having to plot the two remaining variables (Y and Z). By generating scatterplots (Y vs. Z), we can generate a movie consisting of two frames (normal and disease). By flipping back and forth between the frames (i.e., two spreadsheets), changes in the ratios become immediately apparent (not illustrated, but examples can be generated using the marker data in the software package). The massive change that occurs in a large data set is at first surprising and then somewhat frightening in that it shows how change actually operates in a complexity.

Figure 17 attempts to illustrate a complex change by superimposing control (yellow) and experimental (blue) data sets; overlapping data appear green. Log-log plots, which spread out the ratio data, reveal a rich repertoire of patterns globally and locally. In autism, the points move inward, outward, or stay the same.

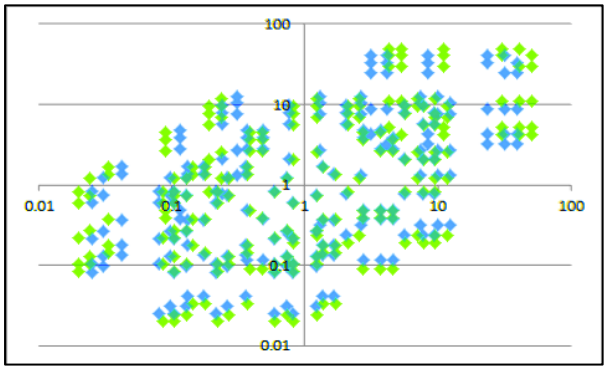

Figure 17. Key: Autism (blue), Normal (yellow), overlap (green). Compared to the normal, autism is characterized largely by a contraction of the point set (inward movement – yellow to blue). However, examples of an outer movement and of no movement (complete green squares - overlap of blue and yellow) can be readily identified in an animation.

Complexity Theory – The Brain

The fact that the same markers can be shared by different diseases – despite their six variable complexities - prompts a closer look at the local and global patterns of disease in the brain. Since pictures often tell the story most effectively, we will use them here to explore the relationship of complexity theory to the brain.

Schizophrenia appears to be the most intrusive disorder of the brain in that it involves the largest number of parts and connections (Figure 18). See also Bolender, 2011.

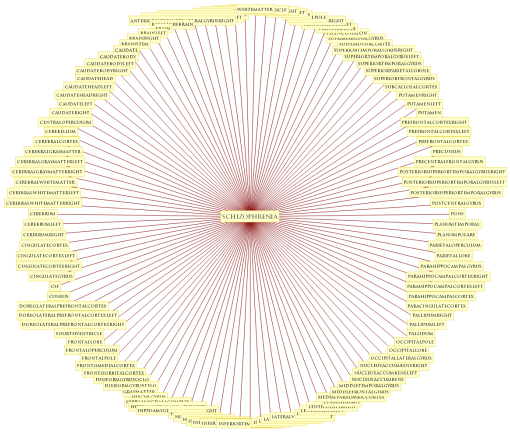

Figure 18. Schizophrenia involves at least 123 parts of the brain. Enlarge as needed.

By adding other diseases to the plot of Figure 18, we can see the relationship of parts to connections for 14 different diseases of the diagnosis database (Figure 19). Schizophrenia remains dominant with its 123 parts and connections, but notice that it shares many of its parts (~30%) with the other diseases. The wholly unexpected finding is that the parts and connections of 6 individually recognized diseases are identical to those of schizophrenia. What does this mean? Is schizophrenia releasing diseases as spinoffs or is schizophrenia an aggregation of many different diseases?

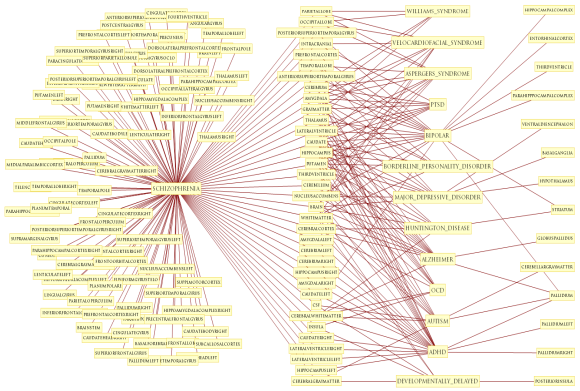

Figure 19. Diseases of the brain share many similar parts and connections. Enlarge the image to view details.

There is more (Figure 20). When we plot just three diseases (schizophrenia, bipolar disorder, and ADHD), the complex relationship of one disease to another becomes clear. Bipolar disorder and ADHD together appear as a distinct subset of schizophrenia in that they share 80% of the same parts and connections. Moreover, bipolar disorder and ADHD appear related in that together they share roughly 25% of the same parts and connections.

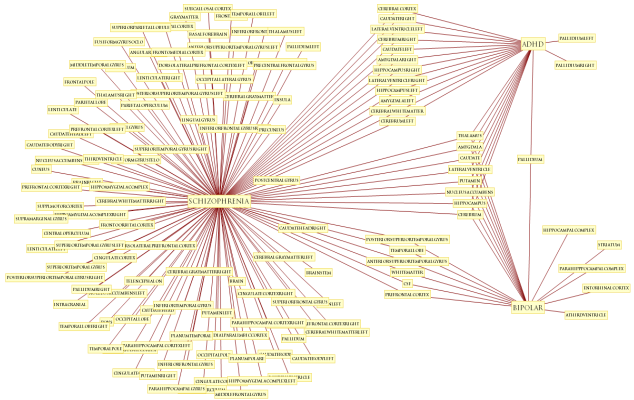

Figure 20. ADHD and bipolar disorder share many identical parts and connections with schizophrenia, as well as with each other. Enlarge as needed.

This pattern of a close relationship between diseases (Figure 20) persists for many other combinations thereof. Figure 21, for example, shows the relationship between bipolar disorder and Alzheimer disease. These two diseases share 9 out of 25 (36%) parts and connections.

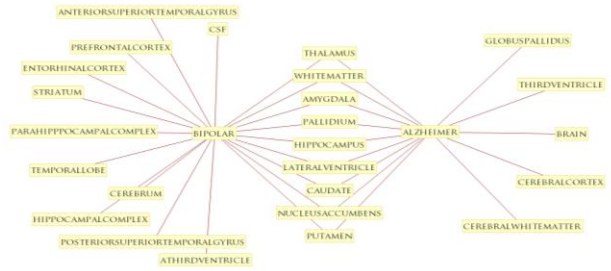

Figure 21. Bipolar disorder and Alzheimer disease share similar parts and connections. This graphic shows that they share a central module consisting of 9 parts.

Perhaps the most interesting part of the story is that different disorders of the brain share not only parts and connections, but also identical mathematical markers. In addition to capturing phenotypes, these markers begin to explain how normal and abnormal brains are constructed hierarchically as complexities based on well-defined relationships of parts to connections. The pattern of connectivity begins with triplets (three connected parts) and continues as triplets combine to produce higher order complexities (modules to disorders). As we move throughout these hierarchical levels, complexity remains a function of the parts and connections, which we can fold or unfold to change their sizes, names, locations, properties, and mathematical markers. The common thread that runs through everything is the universal connectivity of the parts.

Starting with our working model for a general theory of biological complexity (Bolender, 2011), we can extend it to include a theory of disease specific to the human brain (Table 2).

Table 2. Complexity Theory of Disease (Human Brain).

Working Theory of Disease (Human Brain) 1.0
• **Mathematical markers, which consist of three parts and three values expressed as a ratio, define a basic unit of order in biology.** • **The complexity of a disease can be captured with mathematical markers.** • **Mathematical markers – unique to disease – are conserved in that the same markers can appear in different diseases.** • **A disease can display a modular structure, wherein the modules may include individual triplets (mathematical markers) or collections of parts and connections largely defining other diseases.**

DISCUSSION

The report offers examples of how technology helps us to operate comfortably within the realm of biological complexity. In turn, this new found accommodation predicts a series of potentially disruptive events wherein rules, directions, and expectations can all change - perhaps sooner than we might imagine. By reducing many types of biological data to common denominators (volume, surface, length, or number), creating dimensionless ratios, assembling ratios into mathematical markers, and storing everything in a common database, we are well on our way to developing a mathematical clone of biology. By reconnecting what was previously disconnected, we now get to play the complexity game – in earnest.

Mathematical Markers

Mathematical markers – in nearly inexhaustible supply – offer ready access to biological complexity. A disease displaying abnormal patterns can be expected to carry millions of such markers, occurring throughout the biological hierarchy. Simply tap into a connectivity network and start asking questions. If we know the rules, we can play the game. By applying the same data format to all parts and all connections, mathematical markers effectively generalize and integrate data across all parts of an organism. This obeys a rule of complexity theory which states that all parts – large and small - are connected qualitatively and quantitatively.

The welcoming nature of this new technology becomes obvious in that it does not discard our current technology based on reductionism, but rather it takes reductionism to a new and more informative level. In the basic and clinical sciences, complexity cannot exist without reductionism because it performs the critical task of producing the raw data needed to build and populate a complexity.

But, do these mathematical markers actually exist in nature or are they simply the result of forming permutations? Recall that biological parts were

used originally to generate data pairs and data pairs data triplets (Bolender, 2010, 2011). Two data pairs can combine to form a triplet when two of the parts share identical names and decimal repertoire values. To demonstrate the existence of triplets, the data of 67 publications (from the IBVD) were used to generate 37,950 data pairs of which 32,138 (84.7%) formed triplets. In short, triplets actually exist. The missing 15.3% may be attributed to parts in transition or to sample sizes with too few parts to detect triplets. Notice that the plot of Figure 22 suggests that papers with larger data sets tend to produce a higher percentage of triplets.

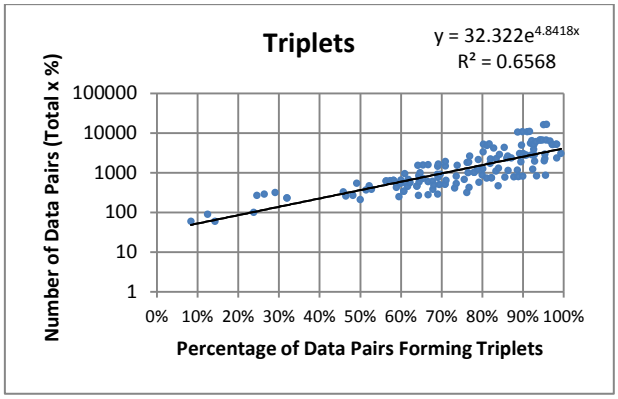

Figure 22. The range of data pairs forming triplets (per paper) extends from 8% to 99%.

The data – in the form of mathematical markers - tell the story most convincingly. Run the software and explore the markers databases. Begin with the normal + disease data set (spreadsheet or database), sort the table on the mathematical markers column, and scroll through the rows. Notice the remarkably long runs of duplicate markers revealing the preferences of biology in creating and conserving specific patterns of parts and connections. Check the citation nu column to confirm that the duplicate markers come from several different papers. Compare the original published values (volumes) to the derived ratios (dimensionless), noting the range of each. One is broad (volumes) the other narrow (ratios).

Mathematical markers quantify disease in terms of parts and connections, but to what advantage? Notice that figures 12, 13, 15, 19, 20, and 21 show different disorders displaying similar patterns, which can be explained - in part - by markers serving as interchangeable modules. This raises an interesting question. If, as these figures suggest, we can unfold the complexity of a disease, can we also unfold the complexity of its prevention and treatment? Such questions quickly snap into view, particularly if we take a hard look at disease as a complexity. Instead of studying hundreds of different brain disorders one by one, why not - since they have so much in common - study them all together as a single group? Might common etiologies lead to common treatments? The results given in Figures 16 and 19, for example, would seem to offer a compelling argument in favor of a global approach to understanding and treating disease.

Diagnosis

We are often reminded that the best and most effective approach to managing disease is prevention, which typically requires early detection and life-style changes. The MRI database suggests that a diagnosis based on mathematical markers can provide not only the starting (normal) and end points (disease), but also points in between. These intermediate points could be highly effective in alerting us to a threatening condition at a time when intervention might be most helpful. In effect, we can use this new technology to create a feedback loop that keeps us informed of our fitness over time.

Consider, if you will, a future scenario based on mathematical markers. At the completion of a MRI head scan, we would read our diagnosis and view a list of options and recommendations. In this case, we elected the scan not to identify the presence of a disorder, but rather to evaluate our overall health and to consider ways of avoiding potential problems. To become an alternative – or adjunct - to the traditional physical exam, a MRI head scan would have to detect changes in the brain produced by conditions existing throughout the body. Indeed, this may be entirely possible. Recent MRI studies report that diseases in the periphery can in fact induce changes in the brain (Clarence et al., 1999, Pérez-Dueñas et al., 2006, Borson et al., 2008, Nagai et al., 2010, Agostini et al., 2012). Perhaps, a diagnosis database

populated with billions - or trillions - of mathematical markers will define a new and enormously powerful health care tool, one that operates effectively at the level of our individual phenotypes.

Developing new technologies for diagnosing disease is becoming a high priority item as evidenced by the announcement of a recent X Prize (Qualcomm Tricorder X Prize). The sponsors want a handheld device that can diagnose disease in an individual automatically – basically a gadget equivalent to the medical tricorder of Star Trek fame. The stated purpose of the initiative is to inject new and disruptive technologies into the health care system. The prize will be awarded to the team that can identify correctly a relatively modest number of disorders (15) – none of which occur in the brain. The most interesting read in the guidelines document (a .pdf document) is that they reveal the names of the 15 diseases that will be on the final test. This appears a curious strategy in that they are asking for a suite of biological markers operating in a reductionist setting even though complexity is most likely to be the principal disrupting agent going forward. Moreover, supplying the answers to the test exposes the outcome of the contest to a worst case scenario. Allow me to explain.

If individual biological markers overlap diseases in ways similar to those seen for mathematical markers in the brain (e.g., Figures 13 and 15) or if the biological markers maintain their track record of being consistently unreliable (Ioannidis and Panagiotou, 2011), then one marker per disease is not likely to be enough to get the job done. Moreover, the teams must deal with the thorny problems of false positives and ambiguous concentrations (Bolender, 2001-2011). Given the obvious pitfalls of a reductionist approach, why not simply solve the X Prize as a complexity?

How might this be accomplished? We would start with the data. Question: What are the five most readily available sources of patient data? Blood, sweat, tears, saliva, and urine would all seem to qualify. If each of these fluids can supply 20 detectable parts (e.g., ions, molecules, cells) then we get 34,200 mathematical markers for each of the 15 specified diseases of the X Prize with a total of 513,000. Chips (e.g., Lab-on-a-chip, IBM Research) and standard laboratory devices can be used to assay the fluids, a handheld computer to generate the markers, and a Wi-Fi connection to send the markers to the cloud for matching the unknowns (patients with one of 15 possible diseases) to the 513,000 standards described above. Now consider the final exam of the X Prize. If one team has 15 markers and the other 513,000, the odds of complexity winning would be roughly 34,200 to 1.

Unified Theory of Diagnosis

The usual incentive for proposing a unified theory is to suggest that a single solution applies to a considerable number of different problems. In disease, where parts and connections undergo changes, diagnosis depends on detecting these events either directly as markers or indirectly as symptoms. Since we now know that we can use markers to diagnose disorders of the brain (Examples 1 to 4) and that triplets occur throughout the organism (Bolender, 2011), mathematical markers become a candidate solution. A unified theory allows us to generalize diagnosis at the level of the organism (Table 2).

Table 2. The unified theory of diagnosis proposes a single data model – based on triplets - for all diseases.

Unified Theory of Clinical Diagnosis 1.0
The theory proposes that: • **Data sources can include the basic and clinical sciences and all other sources of data.** • **Triplets, consisting of three biological parts, can be derived from volume, surface, length and number data by forming all possible combinations.** • **Mathematical markers can be formed by concatenating the parts and connections (ratios) of triplets.** • **A single database table can include mathematical markers coming from all parts and all diseases.** • **A disease can be diagnosed by running its markers against known standards (identified markers) in the diagnosis database.** • **Specific risk factors related to life-style, employment, drugs, treatment protocols, et cetera can be assigned to individuals based on their phenotypes – as defined with mathematical markers.** • **Mathematical markers can create a complex data set well-suited to preventative, diagnostic, and predictive goals. As such they suggest a general solution to the problem of disease as a complexity.**

The unified theory of diagnosis hypothesizes that mathematical markers can diagnose all diseases provided that sampling is unbiased, that parts can be identified and estimated accurately, that changes in parts and connections occur, and that parts are available in sufficiently large numbers. To test the theory, we can populate the current diagnosis database with data coming from all the other parts of the body and then run the markers of unknown diseases against the known standards. Generalizing diagnosis, of course, moves us one step closer toward generalizing disease.

Disease

The power of science derives from its ability to generalize large and complex bodies of information. In turn, generalizations drive progress by being translated into new theory structure the purpose of which is to encourage new levels of understanding.

What happens when we apply this formula to human disease? We quickly gain a radically different perspective. By unfolding diseases into their component parts and then refolding the parts into mathematical markers, it becomes apparent that all the diseases considered thus far conform to the same design strategy – a stoichiometric arrangement of parts and connections (Bolender, 2001-11). Such a finding should not come as a surprise in that an orderly arrangement of parts is a design principle fundamental to biology, chemistry, and physics.

But why does the brain display such a broad range of disorders based on a similar underlying design? What is the advantage? Are we simply looking at the consequences of damage and unfortunate mistakes, or are these disorders somehow related to the behavior of an complex adaptive system? We know, for example, that a prime direction of living systems is to evolve into new and more successful configurations. This is achieved by reshuffling old or introducing new parts and connections, enabling new properties to emerge.

How do disorders of the brain fit into this overarching picture of change? Although we normally associate brain disorders with negative outcomes, disorders creating geniuses, creative artists, and savants are among the most dazzling examples of an innovating brain. Might some or several brain disorders simply be the result of the brain trying to evolve and become more successful? The question need not remain an academic one because the mathematical markers of geniuses, for example, can be assembled from MRI head scans and run against the markers of the current disease database. An interesting result might encourage a young and adventurous scientist to follow a parade of markers all the way back to the genome. By keeping everything well-connected, complexity theory actively encourages such activity.

Concluding Comments

The report accomplishes two things. It introduces the reader to the application of complexity theory in biology and provides a support structure for the theory in the form of a software package. Complexity is rapidly becoming a new frontier in biology because it represents a primary source of new information that can drive discovery, innovation, and productivity very effectively. This new direction signals a maturity in our understanding in that we now know that it takes a complexity to solve a complexity.

With all the major parts in place, all we need are the right catalysts to activate the complexity machine. By sheer chance, I discovered the Internet Brain Volume Database sitting in a collection of largely molecular biology databases. As a result, we now have almost a million mathematical markers, a general method for diagnosing disease, working theories for complexity, diagnosis, and disease, and even an ad hoc solution to a coveted X Prize. For me, this database became the catalyst that turned otherwise fanciful ideas into reality by providing ready access to published data. Since most biomedical publications continue to exist behind paywalls, this was an extraordinary find. Just imagine how much more exciting and productive science would become if we had open access to published data and could browse them freely in databases.

REFERENCES

Agostini A., Benuzzi F., Filippini N., Bertani A., Scarcelli A., Farinelli V., Marchetta C., Calabrese C., Rizzello F., Gionchetti P., Ercolani M., Campieri M. and P. Nichelli. New insights into the brain involvement in patients with Crohn's disease: a voxel-based morphometry study. Neurogastroenterol Motil. 2012 Sep 23. doi: 10.1111/nmo.12017.

Bolender, R. P. 2001a Enterprise Biology Software I. Research (2001) In: Enterprise Biology Software, Version 1.0 © 2001 Robert P. Bolender

Bolender, R. P. 2002 Enterprise Biology Software III. Research (2002) In: Enterprise Biology Software, Version 2.0 © 2002 Robert P. Bolender

Bolender, R. P. 2003 Enterprise Biology Software IV. Research (2003) In: Enterprise Biology Software, Version 3.0 © 2003 Robert P. Bolender

Bolender, R. P. 2004 Enterprise Biology Software V. Research (2004) In: Enterprise Biology Software, Version 4.0 © 2004 Robert P. Bolender

Bolender, R. P. 2005 Enterprise Biology Software VI. Research (2005) In: Enterprise Biology Software, Version 5.0 © 2005 Robert P. Bolender

Bolender, R. P. 2006 Enterprise Biology Software VII. Research (2006) In: Enterprise Biology Software, Version 6.0 © 2006 Robert P. Bolender

Bolender, R. P. 2007 Enterprise Biology Software VIII. Research (2007) In: Enterprise Biology Software, Version 7.0 © 2007 Robert P. Bolender

Bolender, R. P. 2008 Enterprise Biology Software IX. Research (2008) In: Enterprise Biology Software, Version 8.0 © 2008 Robert P. Bolender

Bolender, R. P. 2009 Enterprise Biology Software X. Research (2009) In: Enterprise Biology Software, Version 9.0 © 2009 Robert P. Bolender

Bolender, R. P. 2010 Enterprise Biology Software XI. Research (2010) In: Enterprise Biology Software, Version 10.0 © 2010 Robert P. Bolender

Bolender, R. P. 2011 Enterprise Biology Software XII. Research (2011) In: Enterprise Biology Software, Version 11.0 © 2011 Robert P. Bolender

Borson S., Scanlan J., Friedman S., Zuhr E., Fields J., Aylward E., Mahurin R., Richards T., Anzai Y., Yukawa M., and S. Yeh. Modeling the impact of COPD on the brain. Int J Chron Obstruct Pulmon Dis. 2008 September; 3(3): 429–434.

Clarence L., Edwards S., Gong Q., Roberts N. and L.D. Blumhardt. Three dimensional MRI estimates of brain and spinal cord atrophy in multiple sclerosis. J Neurol Neurosurg Psychiatry 1999;66:323-330 doi:10.1136/jnnp.66.3.323.

Ioannidis J.P. and O.A. Panagiotou. Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. JAMA. 2011; 305 (21): 2200-10.

Keller S.S. and N. Roberts. Measurement of brain volume using MRI: software, techniques, choices and prerequisites. J Anthropol Sci. 2009;87:127-51.

Kennedy D.N., Hodge S.M., Gao Y., Frazier J.A., and C. Haselgrove. The internet brain volume database: a public resource for storage and retrieval of volumetric data. Neuroinformatics. 2012 Apr;10(2):129-40.

Nagai M., Hoshide S., and K. Kario. The insular cortex and cardiovascular system: a new insight into the brain-heart axis. J Am Soc Hypertens. 2010 Jul-Aug;4(4):174-82.

Qualcomm Tricorder X PRIZE: Healthcare in the palm of your hand. http://www.qualcommtricorder xprize.org/competition-details/overview

Pérez-Dueñas B., Pujol J., Soriano-Mas C., Ortiz H., Artuch R., Vilaseca M.A., Campistol J. Global and regional volume changes in the brains of patients with phenylketonuria. Neurology, Apr 2006; 66: 1074 - 1078.

APPENDIX

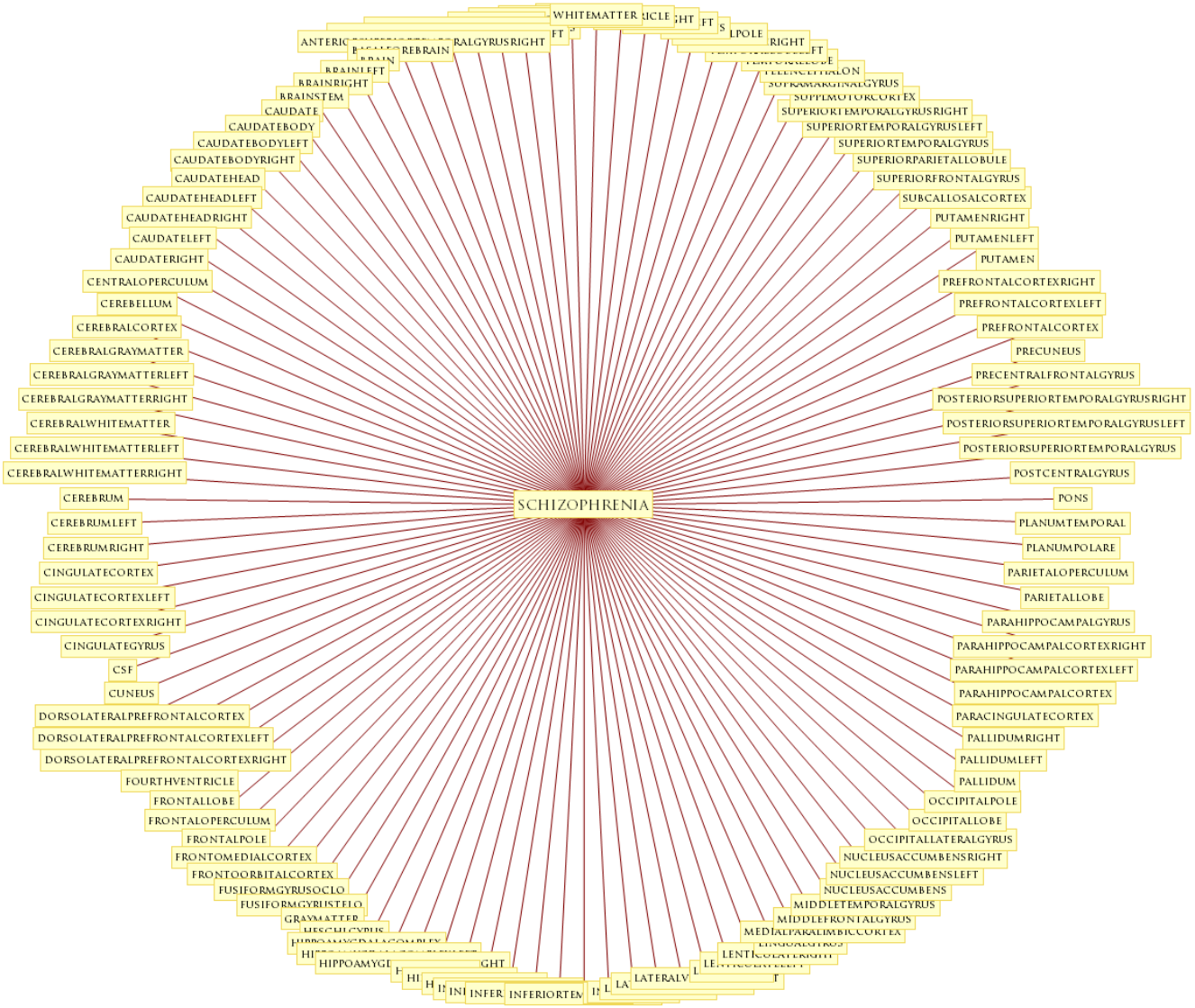

Figure 18. Schizophrenia involves at least 123 parts of the brain. Enlarge as needed.

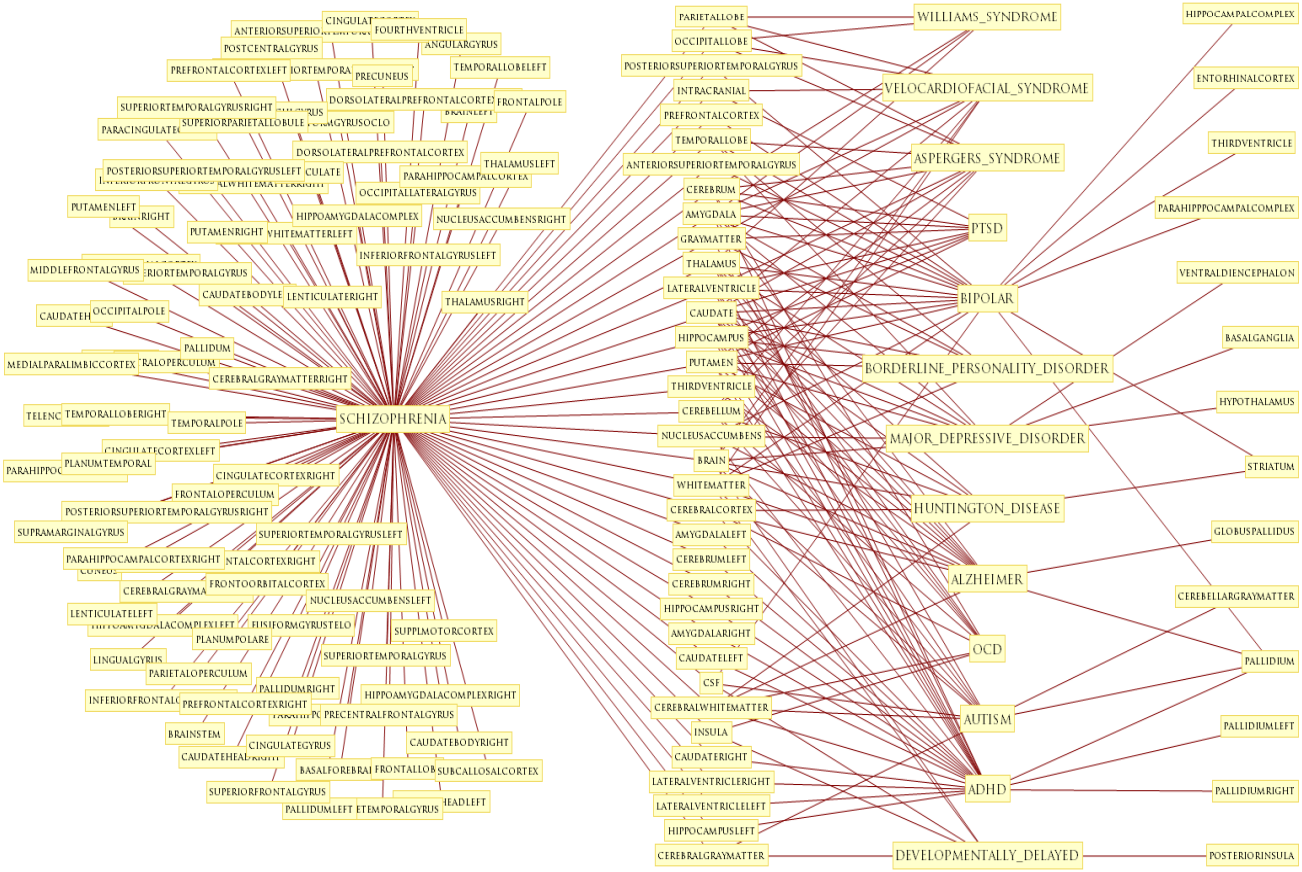

Figure 19. Diseases of the brain share many similar parts and connections. Enlarge the image to view details.

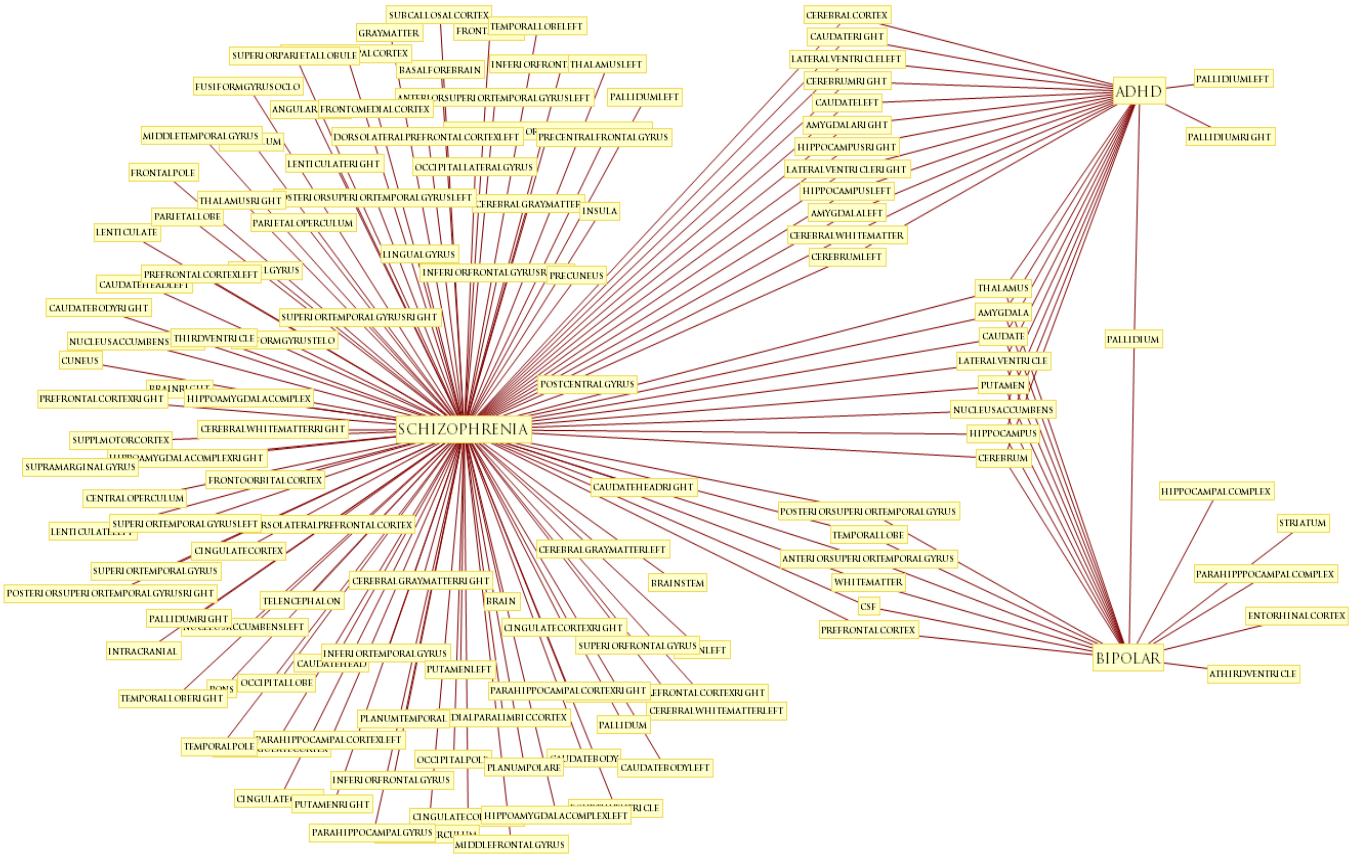

Figure 20. ADHD and bipolar disorder share many identical parts and connections with schizophrenia, as well as with each other. Enlarge as needed.