Are you nothing but genes or neurons?

Sunny Y. Auyang

All complex systems are complex, but some are more complex than others are. Biological systems are generally more complex than physical systems. How do biologists tackle complex systems? In this talk, we will consider two biological systems, the genome and the brain. Scientists know much about them, but much more remains unknown. Ignorance breeds philosophical speculation. Reductionism makes a strong showing here, as it does in other frontier sciences where large gaps remain in our understanding. I will show that reductionism and its claims have no bases in actual scientific research and results. The Human Genome Project will serve as a case in point.

What is new in those "astonishing hypotheses?"

Materialism, the doctrine that everything is physical and the world contains no soul substance or vital spirit, has a long root in Western thinking that reaches back to the pre-Socratics. In the eighteenth century, when people worried strongly about the immortal soul, Joseph Priestley argued, "the powers of sensation or perception, and thought, as belonging to man, have never been found but in conjunction with certain organized system of matter... [W]e could not but have concluded, that in man thought is a property of the nervous system, or rather of the brain."[1] Priestley did no lack supporters in the debate about mind and matter. Opposition weakened with secularization and scientific progress. At the end of the nineteenth century, the young Sigmund Freud reflected the intellectual climate of his time in the opening declaration of his A Project for Scientific Psychology: "The intention is to furnish a psychology that shall be a natural science; that is, to represent psychical processes as quantitatively determinate state of specifiable material particles."

Strangely, by the end of the twentieth century, materialism, the old commonplace, becomes a bold and novel proposition. Looking at the scholarly literature, one cannot help sharing the puzzle of Noam Chomsky, who observed: "Every year or two a book appears by some distinguished scientist with the 'startling conclusion' or 'astonishing hypothesis' that thought in humans 'is a property of the nervous system, or rather of the brain.'"[2] What has gone wrong with our times?

Francis Crick, who with James Watson discovered the DNA structure, has turned to neuroscience. In a popular book, he advanced what he called an "astonishing hypothesis." "You, your joys and your sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behavior of a vast assembly of nerve cells and

their associated molecules." In short: "You're nothing but a pack of neurons."[3] *Nothing but*, that is the astonishing element that our age adds to the time-honored materialism.

The claim that a thing is *nothing but* something else is the crux of ideological reductionism, an influential philosophical doctrine in this century, although its prestige is being increasingly challenged, especially in the study of complex systems. I prefer to call it *nothing-butism*, to avoid confusion with the more productive meanings of reduction as analysis or decomposition. Nothing-butism asserts that a complex system is no more than its parts, so that we understand the system exhaustively once we understand how its parts behave. You are nothing but a pack of neurons, therefore scientists can predict all your joys and sorrows once they know how neurons work.

Eliminativism is the name of nothing-butism in the philosophy of mind. It denies the possibility that high-level properties that we ordinarily describe in mental terms can emerge from the complex large-scale organization of neurons. Banning emergence, mental properties become nothing but neural properties. The mental vocabulary that we use in our daily life should either be identified with neural terminology or be eliminated from our "scientific" discourse.

For a while, nothing-butism seemed to make sense. In the 1950s, cells specialized for certain functions are discovered. For instance, the frog's retina contains cells that are sensitive to the movement of small objects in a wide range of illumination. These cells are called "bug detectors," although frogs do not differentiate bugs from other small moving objects. David Marr recalled what happened: "People were able to determine the functions of single elements of the brain. There seemed no reason why the reductionist approach could not be taken all the way." In their excitement, people speculated that vision could be explained in terms of the excitation of specialized neurons, for example the fabled "grandmother neuron" that fires only when one's grandmother comes into sight. The mind-brain identity thesis advocating "pain is nothing but c-fiber firing" appeared around this time. Although the philosophy survives today, the scientific excitement died quickly as scientists realized that the behaviors of individual neurons give no clue to how they contribute to vision as a whole.

Neurons sensitive only to a specific feature exist. Nevertheless, their behaviors are influenced by mental behaviors such as an animal's attention. For instance, a direction-sensitive neuron fires only if its receptive field contains an object moving from right to left. Experimentalists use microelectrodes to measure the activity of a single neuron in a monkey trained to attend to only one of two objects in its view.   The two objects move horizontally in opposite directions. If the direction-sensitive neuron operates autonomously, it should always respond in the same way, because one of the two objects moves in its preferred direction. This is not what happens. The neuron responds strongly only if the attended-to object moved from in its preferred direction. Its response is markedly weaker if the ignored object moved its preferred direction. Attention and other global mental properties made a big difference on the behavior of individual neurons. Similar results obtain in many other experiments. They refute nothing-butism, which, by dismissing mental effects such as attention, makes a total mystery of why the direction-sensitive neuron responds differently to the same stimuli.

Are you controlled by selfish genes?

The most famous version of nothing-butism in genetics is Richard Dawkins's doctrine of the "selfish gene" that features genes each fighting for its own benefits. The organism as a whole is nothing but a dead vehicle controlled by the selfish genes. Curiously, the genes are defined not by their molecular structures but by their functions, as a gene for intelligence or a gene for sexual preference. The gene for homosexuality fends for itself regardless of other genes, the structures of the genome, and the conditions of the organism. The selfish gene story is popular because it is easy to understand. Easiness, however, does not imply correctness.

Scientists occasionally find genes for specific functions, but these are exceptions rather than the rule. The most promising cases for selfish genes are the genes for diseases; it is always easier for a single gene to foul things up and lonely culprits are easiest for researchers to track down. Even so, they are in the minority. Of some 5,000 known genetic disorders, only a fraction are attributable to single genes, for examples cystic fibrosis and Huntington's disease. Even in these cases, researchers warn that the single-gene mutants are informative only when pursued within the framework of gene linkage.

For more complex diseases, enthusiastic genetic announcements often fall apart under scrutiny. Such is the fate of the gene for schizophrenia and the gene for manic depression. In the overview of a special issue on genetic disease of the journal Current Opinions in Genetics and Development, the editors wrote: "Complexity is now the watchword of the day in disease genetics. . . . Even the apparent simplicity of single-gene disorders is being clouded by the specter of modifier genes that can influence disease susceptibility, severity, or progression."[4]

It takes only a malfunctioning part to disable a car, but many things must work to make the car run properly. Traits for normal behaviors are more complicated than diseases. Chances are much smaller that there are specific genes for traits such as intelligence or sexual preference. The gay gene or the gene for male homosexuality, announced with fanfare in 1993, turns out to be a house of cards. It is almost the rule that many genes contribute to a trait, and a gene is expressed in many functions. Gene expression is a complicated process involving many epigenetic and environmental factors. All these are ignored by the story of the selfish gene.

The gene for homosexuality and the c-fiber for pain share the central dogma of nothing-butism, which allows only one level of description for complex systems. Sexual preference and pain are traits of the organism, which belongs to a higher level of organization than DNA molecules or neurons. Nothing-butism tries to eliminate the level of organisms in favor of the level of genes or neurons. What it achieves is nothing but a confusion of properties on two distinct levels.

Atomization, construction, synthetic analysis

Atomization is different from analysis, which many scientists call reduction and from which I take pain to distinguish reductionism. To atomize a system is to demolish it into an aggregate of elements. Nothing-butism takes the extreme atomistic view that once we know the properties of individual elements, we can combine them in various ways without any idea of the whole. All

we need are concepts for the elements and their relations, concepts of the system are superfluous. This is what it means by the system is *nothing but* the elements. It is a purely bottom-up approach, where one constructs from given elements.

In contrast to atomize, to analyze a system is not to analyze it away. The aim of analysis is to understand the system. To emphasize the point I call it synthetic analysis. Suppose we analyze a geometric whole into two parts by drawing the boundary *R*. *R* not merely differentiates but simultaneously unites the two parts that it delimits, and the parts are always understood as parts of a whole. Further analysis reveals finer structures of the whole by drawing more distinctions and delineating more intrinsically related parts. In theoretical science, the "boundaries" that cut up the whole into "parts" are concepts. In Plato's metaphor, we conceptually carve nature at its joints. The trick, as Plato said, is to observe the natural joins, not to mangle it up like an unskilled butcher.

The synthetic analysis of complex systems differs from atomization and bottom-up construction in several major ways. First, it keeps in view at least two distinct levels of organization, which it describes in different concepts and tries to connect. Unlike atomization, which destroys the systemic level, synthetic analysis never loses sight of the whole system, even when it examines the parts. Second, it is a round trip from the whole to the parts and back. Analysis proceeds from the top down, but the analytic results must be eventually synthesized to obtain desired answers for complex systems. The final synthesis is not a purely bottom-up construction because it is guided by the original analysis, so that it is not overwhelmed by the combinatorial explosion. Third, it is not bound by pre-fabricated parts. Scientists have the freedom to choose the appropriate depth of analysis for the explanation of certain system behaviors.

Consider for example genetics and molecular biology. Nothing-butism demands them to start with a bunch of nucleotides and try to piece them together. In contrast, scientists start with the notion of inheritance of the traits of organisms. They analyze organisms, identify chromosomes, analyze them to find genes, and then analyze molecular structures of the genes to find their nucleotide base sequences. It is important that they do not stop with the base sequences. They proceed to examine how the sequences code for amino acids, which combine to form proteins, which facilitate metabolism and keep the organism alive. The recent Human Genome Project is a good example of analysis followed by synthesizing the results to understand organisms. Notice that even in the most detailed analysis where one talks about the sequence of nucleotide bases, the bases are never isolated atoms but occur in the context of the gene and genome.

The genome and the brain

Almost all organisms share the same genetic code. It is made up of four kinds of nucleotide bases: adenine, guanine, cytosine, and thymine; A, G, C, T. The nucleotides bind to each other, forming a long chain, a DNA molecule. Usually the DNA molecules coil in a double helix. When they uncoil, a long chain of DNA, AGTCCCAAGT . . . serves as a template for reproduction and transcription. When a gene is expressed, a segment of DNA is transcribed to an RNA molecule. In complex organisms, most of the DNA does not code for anything, and they form segments called introns. These introns are cut out of the RNA to form a shorter

message-RNA. The message RNA is then translated into a protein. Three consecutive nucleotide bases of the mRNA, e.g. AAG, constitute a codon that specifies an amino acid. There are twenty kinds of amino acid, which are the building blocks of proteins.

The genome of any organism, even a worm, is a very complex system. *C. elegans*, a roundworm, is the first multicellular animal whose genome has been completely mapped. It has only about 900 cells, but it has some twenty thousand genes. Humans are more complicated, but not by many orders of magnitude. The human genome consists of about 3 billion nucleotide bases divided into 23 chromosomes. The 3 billion nucleotides are grouped into around 65,000 to 80,000 genes. The sizes of the genes vary greatly, but on the average, a gene contains about 10 to 15 kilobases. Only 3% of the DNA codes for proteins. Other portions regulate the expression of the coding genes. However, the vast majority of the DNA bases have no clear functions. Probably they are junk.

DNA and proteins are large biological molecules, polymers. They distinguish themselves from other polymers by their complexity. Many polymers are huge but simple. For example, glycogen contains thousands of glucose units but is simple, because all units are the same and are connected in the same way into a one-dimensional chain. The whole is simply the monotonous repetition of its parts. It has very low information-content complexity. A system's degree of information-content complexity is measured by the length in bits of the smallest program capable of specifying the system to a computer.

In contrast, there are four kinds of base units for the nucleic acids and twenty kinds of base units for proteins. The succession of different bases in the chain creates endless variety. The variety ensures high information-content complexity; to specify a particular sequence of DNA, we need almost to repeat all its nucleotides.

Variety itself does not guarantee interest. Significance depends on the context where variety occurs. The molecules in a glass of water have velocities that vary infinitely. However, they sit in a system where their variety can be averaged over to yield gross regularities. In contrast, DNAs and proteins operate in a system, the organism, where specific details of their individual structures can have singular consequences. The demand for details and specificity enhanced biological complexity.

The brain, by which I mean the central nervous system, is even more complex than the genome. There are more nerve cells, neurons, than nucleotides. More important, there are far more kinds of neurons, each neuron has more complicated structures and dynamics than a nucleotide, and the relations among the neurons are far more intricate. The nucleotides form a linear chain, so that each base is related only to its two neighbors with unaltered bond strength. In contrast, each neuron is connected to about a thousand other neurons, forming a complicated web, and the connection strengths are not fixed but change in time. In view of the variety of neurons and synapses, the multiple connections, and the ever-changing connective strengths, I think neuroscientists are not exaggerating when they say that the brain is the most complex system that we know.

Synthetic analysis of the genome and the brain

Actual approaches in brain or genome research are synthetic analytic. They firmly retain the level of the brain or genome when they delve into neurons or genes. To analyze a system, one must first have some understanding of the whole. This understanding constitutes the synthetic framework in which scientists raise significant questions for analysis to answer. We have seen some examples of synthetic frameworks this morning. Hydrodynamics provides the synthetic framework for the microanalysis of fluids. The state space provides a theoretical framework for chaotic systems. The synthetic frameworks in biology and neuroscience are not mathematical, but they are no less important. Models in psychology and our commonsense concepts of mental properties, including perception, thinking, and consciousness, constitute the framework for the analysis of the brain. The theories of biological evolution, heritability, and genetics provide the synthetic framework for the molecular analysis of the genome.

The most prevalent general conception that represents the system in analysis is the notion of function. Genes, for example, serve the function of inheritance. Similarly, neuroscientists distinguish various regions of the brain according to their functions, as the regions for vision or language. Functional concepts imply a larger system in which the service or contribution of an entity is significant, as the functions of various organs in an organism. Even if the system is tacit in some theoretical models, its demands hide in the notion of an entity's functions.

Vision and sexuality are traits of organisms. Functional definitions such as visual neurons or gay genes demand explanations for neural or genetic contributions to organism behaviors. Usually, the contributions are very complicated, because visual experience or sexual preference involves many intricately tangled factors. With its notion of a neuron for seeing one's grandmother or a gene for homosexuality, nothing- butism covers up the complicated phenomena by collapsing the level of the organism. It trivializes the notion of function by simply identifying the organismic trait to be explained with a fictitious neuron or gene. Its trick is a generalization of the homuncular fallacy.

The homuncular fallacy is a common mistake in discussions on mind. It explains how you recognize your grandmother by positing a little man, a homunculus in your head, who recognizes the grandmother. It is fallacious because it claims to explain recognition, but it explains nothing. It merely shifts the cognitive ability from the man to the little man, therefore its "explanation" presupposes recognition, the trait to claims to have explained. Nothing-butism commits the same fallacy. The grandmother neuron plays the same role as the fictitious homunculus. It covers up complex problems by offering simplistic answers that, when examined, turn out to be vacuous if not false.

Synthetic analysis in the Human Genome Project

The human genome project is arguably the most significant project in biology. For historians and philosophers of science, it has the added advantage of being a coordinated research effort, so that its rationales are publicly debated. Genomic research has tremendous practical ramifications and involves ethical, legal, and social questions. We will leave these questions aside and

consider the project only as a scientific campaign to tackle with a class of complex system, namely, genomes of various organisms.

The major goal of the human genome project is to map the entire human genome, that is, to determine the sequence of all those 3 billions nucleotides and to identify the tens of thousands of genes. This is the flagship of a battle group, whose smaller vassals are responsible for sequencing the genomes of some model bacteria and animals: the bacterium *E. coli*, the single-celled yeast, the worm *C. elegans*, the fruit fly *drosophila*, on which biologists have been working for decades, and the mouse, on which most of our new drugs are tested. The international project broke ground in 1990 and envisioned three five-year plans. Bucking the rule for government projects, it runs ahead of schedule and below budget. Already it has completed the genetic map for yeast and recently, the worm. It expects to finish the sequencing of the human genome by the end of 2003. It has to hurry, because it has stiff competition from the private sector. The privately owned gene factory of Celera, which has come on line a few months ago, vows to sequence the human genome by the end of 2001.

The human DNA sequence, which fills 1,000 books each with 600 pages, analyzes genetics to the smallest biological scale. It is the reductionist dreams come true. What more does one need to understand human genetics? Nothing, nothing-butists claim. Lots more, biologists counter.

From the inception of the project, biologists argued that the sequence alone is worthless. Robert Moyzis, director of the Center for Human Genome Studies at Los Alamos, said: "Many individuals with a physical-science background do not understand that a DNA sequence without a genetic map is nearly useless." David Botstein, Chairman of the Department of Genetics at Stanford, explained why: "the straight sequencing approach was crazy because it ignores biology."[5]

The DNA sequence provides mainly *molecular* knowledge. Its major value lies not in itself but in its contribution to acquiring knowledge of *biological* phenomena that occur on higher levels of organization. This is eloquently expressed in the policy of Celera, which announces that it will make its data on the DNA sequence freely accessible online. Celera is not a philanthropist but a capitalist enterprise out to make money. It figures that big profits lie not in the sequence data *per se* but in making sense of the data, patenting genes, and licensing the manageably packaged information to pharmaceutical companies, which will mine the data to find profitable medical applications.

The Human Genome Project's approach is top down. It starts with the genetic map and the physical map for the genome, which serve as the synthetic framework that gives biological meaning to the more analytic DNA sequence that follows. A genetic map is obtained by studying the genetic compositions of the members of families to find out how various genes are passed on from parents to offspring. Several genes on the same chromosome are often linked and inherited together. Geneticists measure the "genetic distance" between two genes according to the statistical frequency that they are separated by crossing over at reproduction and therefore are not inherited together. 1 cM (centiMorgan) is a unit of genetic distance. It means that the two genes have a 1 percent probability of not being inherited together. Genetic distance does not vary linearly with physical distance on the chromosome. In humans, 1 cM roughly corresponds

to 1 million nucleotide bases. Although crude, the genetic map tells biologists how fragments of DNA sequences are inherited. Without it, all the details of the DNA sequence are useless.

The genetic map gives only the relative "distance" between genes. The physical map fixes these genes to specific positions of the chromosome by using sequence tagged sites. Each sequence tagged site is a short and unique segment of DNA, which serves as a signpost for a location in the chromosome. The 52,000 tagged sites give a resolution of about 60 kb per marker. This is better than the 1 Mb resolution of the genetic map, but it is still crude. The genetic map is like a map of cities and towns, the physical map gives the streets in the towns. Finally, the DNA sequence will give every house address on the streets.

The procedure of the Human Genome Project reveals a synthetic analytic process. It starts from the top and identifies biological goals. Always keeping the genome as a whole in sight, it performs deeper and deeper analysis, going from the genetic map to the physical map to the final DNA sequence.

Synthesis beyond the Human Genome Project

What happens when it reaches the analytic bottom, when we know every one of the 3 billion bases in the human DNA sequence? Only three percent of sequence codes for proteins. The first step to make sense of the sequence is to find those coding segments and identify the protein products they code for. This scientists can do by taking cue from the message-RNAs, and they are well on their way of identifying the coding genes. However, gene identification is far from the end of the story. Each cell in the body contains the whole genome, but most of the genes in it are dormant. Cells being highly differentiated, each cell has its own special needs. Only a small portion of the genes are expressed at any time to produce the specific proteins for a kind of cell. Gene expression is the central scientific problem. Its complexity is overwhelming.

To make sense of DNA sequences, researchers proceed to "functional genomic," which aims to interpret the functions of the expressed DNA sequence on a genomic scale. Here the nothing-butist claim is most pointedly battered. Genes sitting on a DNA molecule are dead things by themselves. They contribute to the life of an organism only when they are expressed in protein products and thereby serve certain functions. Complicated as gene structures are, gene expressions are more complex, for they invariably involve gene interactions. To elucidate the function of a gene, knowledge about the gene's structure is no enough. A gene can have many products, depending on the regulatory processes or on how its message RNA is spliced. Frequently, cellular conditions beyond the genome influence its expression. At what stage of development and under what conditions is a gene active? What are the regulatory and signaling mechanisms that activate it? What is the protein it produces? What kind of cell or tissue does the protein help grow? What are their effects on the organism's physiology? These questions cannot be answered on a gene-by-gene base. Scientists must take the "global view" that accounts for the linkage among genes, the effects of the structures of the chromosome, and other environmental factors.

In anticipation of the need for functional studies, the Human Genome Project includes research on model organisms. To probe the effects of genetic expressions on physiology, scientists must to experiment on the organisms. They knock out a gene or induce a genetic mutation and observe its effects on the organism. Does the organism get a disease or grow a leg on its head? Because scientists cannot experiment on humans, mice and other model organisms come in handy. There is a limit to the usefulness of model organisms. Humans and mice share much genetic makeup, but differences are also great. Discrepancies are enhanced by the fact that a gene's behaviors are influenced by the global conditions of the genome. Because the genomes differ, a mutation of a key gene often produces different effects in humans and mice. For example, when a mutation in the putative "breast cancer gene" kills mice in the embryo, but allows humans to undergo normal growth through much of adulthood.

The complexity of mechanisms that regulate gene expression is apparent even for the single-celled yeast. Yeast has only 6,297 genes. However, the expression for more than 300 genes can be altered by the mutation of a single gene. Yeast synthesizes ethanol from sugar, an activity we utilize in brewing. Its metabolism changes from anaerobic fermentation to aerobic respiration when the sugar concentration in its environment drops. Nothing-butists may tell the story of the gene for fermentation switching off and the gene for respiration switching on. Reality is far more interesting. During the transition, the expressions of almost 30% of yeast genes increase or decrease by two folds. Usually, the expression of a gene is regulated by dozens of proteins. Some proteins act on the gene specifically, others have broader effects; some bind to the DNA all the time, other only briefly. The relative concentrations of the proteins in turn depend on the function of other genes and the cell conditions. In short, higher-level structures exert certain feedback regulations on the activities of genes. In view of such results, and anticipating that the gene activity during human learning or memory will be even more complex, biologists increasingly realize the futility to monitor the activities of single genes alone. A synthetic view of the genome or parts of the genome as a system is indispensable.

As in all complex systems, most genetic complexity lies not in individual parts but in the interaction among the parts. Interaction controls gene expression. Thus systematic methods must be developed to study, simultaneously, the activities of hundreds or thousands of genes. DNA sequencing is not simple, but it appears trivial compared to the complexity in the synthesis of gene expression. Biologists are working very hard to develop experimental techniques for profiling gene functions and expressions systematically. The front-runner now is the microarray. An array contains thousands of identified genes, whose activities can be studied under controlled conditions. Experimental techniques are not enough. The human genome project prides itself for its high throughput. It generates an enormous amount of data, which will grow exponentially with the combinatoric explosion of genetic expressions. To avoid an information overload, there is a dire need of synthetic concepts and methods to mine and organize the data.

Two Big Sciences and two unities of science

The Human Genome Project and its descendents constitute a Big Science. It differs from the other Big Science, elementary particle physics, in three ways. Elementary particle physics is theoretically minded, pursuing phenomena totally detached from our daily business. Genome

study is most practical, pursuing drugs for curing various diseases. Physics is concerned with the simplicity of fundamental physical laws, genome study confronts complexity. Simplicity has a unified form, thus the personal composition of physics is rather homogeneous. Complexity has many facets, thus the effort to tackle complexity of gene expression becomes increasingly inter-disciplinary. It spawns multidisciplinary centers in major universities that draw together chemists, biologists, physicists, and engineers to combat the complexity of gene expression.

The unity of science has been an ideal for centuries. Positivism, with its reductionist bent, advocates an imperial unity where all authority flows from a set of substantive laws and the reduced sciences are annexed if not abolished. Imperial unity seems to suit the Big Science ideal of elementary particle physics. However, it comes under fire recently. The fashion in philosophy of science is the disunity of science. Relativism proclaims an anarchy in which various theories are incommensurate, people subscribing to them cannot communicate rationally, and going from one to another is akin to a religious conversion. Both reductionism and relativism are too simplistic. An empire is not the only form of unity and anarchy is not the only way to plurality. Unity can be achieved as a federation of autonomous states with their own laws agreeing to the general terms of a constitution. The building boom of multidisciplinary centers is an example of the federal unity of science at work. There are still biologists and physicists in the centers for biophysics or physical bioscience, but they are able to communicate with each other and collaborate to achieve a unitary goal. Contrary to relativism, they prove the rationality of science.

Notes

1. Priestley, J. (1777). *Disquisitions Relating to Matter and Spirit*. New York: Garland Pun. Inc. (1976).
2. Chomsky, Language and Nature. *Mind*, 104, 1-61, (1995).
3. Crick, *The Astonishing Hypothesis*. New York: Charles Scribner's Sons (1994, pp. 3, 11).
4. Willard and Davids, "Genetics of disease: Complex genetics, complex diseases." Current Opinion in Genetics and Development 8, 271-3 (1998).
5. In *The Human Genome Project*, ed. by N. G. Cooper, Mill Valley, CA: University Science Books, (1994), p. 71.

Talk presented in the Department of History and Philosophy of Science
University of Sydney
May 1999