The Mismeasure of the Gene

Ruth Hubbard

> Conscious fraud is probably rare in science. It is also not very interesting, for it tells us little about the nature of scientific activity. Liars, if discovered, are excommunicated; scientists declare that their profession has properly policed itself, and they return to work, mythology unimpaired and objectively vindicated. The prevalence of unconscious finagling, on the other hand, suggests a general conclusion about the context of science. For if scientists can be honestly self-deluded . . . then prior prejudice may be found anywhere, even in the basics of measuring bones and toting sums.
> -Stephen Jay Gould, *The Mismeasure of Man.*

Science is an interpretation of nature and, like other forms of interpretation, fits into the cultural framework of its time. I shall illustrate this fact by tracing some of the threads that, in the course of the twentieth century, have led to the notion that genes determine virtually all physical and social characteristics of humans and other animals. Currently, everything about us is "in the genes," offering the hope that, once we learn to read our "genetic blueprint," we will be able to change it and live happily ever after.

The most obvious place to begin this story is with the Czech monk Gregor Mendel who, in the 1860s, developed what have come to be known as Mendel's laws of inheritance. Using pea plants as his experimental objects, Mendel examined the transmission of flower color and of the shape and texture of the seeds to successive generations. He deliberately selected these traits because they are transmitted in an all-or-nothing fashion rather than traits that vary continuously, such as weight or size. After performing large numbers of crosses between plants, which themselves had been shown to breed true, he was able to describe the numerical regularities in the way the traits were passed from parents to successive generations of offspring that have come to be known as Mendel's laws. He, however, did not speculate about what mechanisms might account for the transmission of traits from one generation to the next and merely suggested that it probably involved "factors" within the plants.

That few scientists paid attention to Mendel's paper when it was published in 1865, presumably had to do with the fact that there was no larger context into which to put his observations. This situation had changed dramatically by 1900, when his paper was independently "rediscovered" in three laboratories. By that time, biologists had observed well-defined structures within the cell's nucleus that took up chemical stains and were therefore called chromosomes. They had further noted that, when cells divide, their chromosomes divide also, so that the two daughter cells end up with the same number of chromosomes as were present in the parent

cell. The chromosomes, therefore, were generally accepted as the bearers of heredity and the idea took hold that Mendel's "factors" bore some relationship to them.

In 1905, the Danish botanist Wilhelm Johannsen coined the word gene to lend more concrete reality to Mendel's "factors." At a time when invisible atoms, electrons, and quanta were being accepted into the world of chemistry and physics, biologists had little problem accepting that heredity also was mediated by invisible material particles. Soon, a series of ground-breaking experiments, done mainly with fruit flies and corn (maize), led them to decide that the genes must lie along the chromosomes, like beads on a string, and that when the chromosomes are replicated during cell division, the genes also get copied.

During the first half of the twentieth century, biologists became increasingly interested in exploring the molecular constitution of cells and the ways molecules participate in the metabolism and growth of organisms. They came up with molecular explanations of human diseases known to have hereditary components, such as sickle cell disease and phenylketonuria (PKU), and identified the specific molecules associated with such conditions.

Chemists and biochemists described various biologically important substances, including vitamins and hormones, and characterized their biological functions in chemical terms. In the process, they identified a series of hitherto unknown carbohydrates and fats and also very large and complex proteins, which had previously been thought to be ill-defined aggregates and not discrete molecules at all. It was an exciting period in which chemically oriented biologists spoke of bringing biology to the molecular level. At the same time, they also tried to understand how different chemical components are integrated into the way whole organisms function, writing books with such titles as *The Organism as a Whole* (Loeb 1916), *The Wisdom of the Body* (Cannon 1932), and *Dynamic Aspects of Biochemistry* (Baldwin 1948).

These kinds of explorations led biochemists to identify protein molecules that function as enzymes, others that mediate muscular contraction and relaxation, and yet others that transport oxygen and CO_2 around the body. As part of these kinds of explorations, biochemists came to realize that chromosomes contain both proteins and another type of very large molecule, called DNA, and this raised the question of the chemical nature of genes: are they made of proteins, DNA, or both?

Initially, many biologists favored the idea that DNA forms an inert chromosomal framework to which protein molecules attach themselves as genes. The reason was that, though DNA is a very large molecule, it is made up of only six different components: a type of phosphate, a sugar, and the four so-called bases that are now familiar to us by the abbreviations A, G, C, and T. The naturally occurring proteins, in contrast, contain twenty different subunits (called amino acids), strung together in many different combinations, and come in many different shapes and sizes. It therefore was easier to imagine that different proteins would be the ones to transmit the various traits for which genes are assumed to be responsible.

In the late 1940s and early 1950s, however, experiments done with bacteria and viruses showed that the hereditary material - the gene - consists of DNA. By then, it had become clear that genes are involved in the synthesis of proteins and biologists had concluded that DNA, in fact, specifies the composition of proteins, but the

mechanism by which this happens was an open question. It is, however, crucial to realize that intriguing as this puzzle was, all this time, DNA was looked on as just one of the sorts of molecules, important to the way cells and organisms function.

All this changed in April 1953, when James Watson and Francis Crick proposed their double helical model of the structure of DNA (Watson and Crick 1953). Since then, DNA has come to be considered the most important molecule in biology and "molecular biology" has come to refer exclusively to the biological functions of DNA.

To understand this shift in outlook, it is important to consider the social and political dimensions of how DNA and the double helix came to be propelled into the center of biological interest. Watson has described the discovery of the structure of DNA, from his point of view, in his best-selling memoir *The Double Helix* (Watson 1968). Though it may be hazardous to do so, it is worth speculating how the story of DNA might have unfolded if one of the other two groups of scientists who were trying to elucidate its structure at that time had "won the race." I am referring to the great chemist Linus Pauling and his group at the California Institute of Technology in Pasadena and to Rosalind Franklin and Maurice Wilkins, two experts in X-ray diffraction analysis at Kings' College London.

For one thing, neither of these two groups was racing. They did not even know there was a race. Only Watson and Crick were racing. As for Pauling, he and his colleagues had recently elucidated the structure of the α-helix, a basic structural component of many of the proteins of biological importance. That was an enormous achievement for which Pauling was shortly awarded a Nobel Prize. Before turning to the structure of DNA, Pauling's group had already determined the three-dimensional structure of the bases that compose DNA. It therefore seems reasonable to assume that, had Pauling been the first to describe the full structure of DNA, it would have been exciting, but would have been just another of his many major accomplishments.

By all accounts, at the time Watson and Crick unveiled their DNA model, Rosalind Franklin was close to solving the structure herself. She had been working on it for about two years and, though no one (including Franklin) knew it at the time, Watson and Crick drew heavily on her X-ray measurements and on the structural information she derived from these to come up with the double helix (Maddox 2002). Had Franklin been the one to solve the DNA structure, she would of course have published it, but might not have announced it with great fanfare because that was not her style. The structure in itself was beautiful and people would have been extremely interested, but it might well not have become the biology-shaking event of the century.

In contrast, from the moment Watson and Crick began to think about how to figure out the structure of DNA and long before they had bothered to find out what was known about its chemical composition, they thought of DNA as "the secret of life." Indeed, Watson writes in *The Double Helix* that, even before they had quite clinched their model, Crick rushed into the pub they frequented to announce in a booming voice that they had "found the secret of life" (Watson 1968, 115). And that is how they communicated the news of the structure to their colleagues and mentors, though their note in *Nature* struck the proper objective tone.

What the Watson-Crick model showed (and most people nowadays can find out by

reading the newspapers) is that DNA can be pictured as two spiral ribbons wound in parallel to form a double helix. The four bases - the A's, G's, C's, and T's - are attached to the ribbons at regular intervals and point toward the center of the helix, hence toward one another much like the teeth on a zipper, except that, in DNA, the teeth meet rather than overlap each other. What made the Watson-Crick model so exciting is the fact that, in order to get the bases to fit into the double helix, an A on one ribbon, or strand, must abut on a T on the other, and a C on one must abut on a G on the other.

This geometrical arrangement means that, to copy DNA - the "gene" - the double helix must merely begin to unwind (or, in this metaphor, become unzipped). Each strand can then serve as a template for the synthesis of its partner. As this synthesis progresses, the old strands and their newly formed partners simply zip up to form two identical copies of the original. In other words, the double helical structure itself explains how DNA - the gene - can get copied. The simplicity of this model has had several ideological consequences. One is that the way DNA is copied has been called "self-replication" and DNA has come to be referred to as a "self-replicating" molecule. Of course, it isn't anything of the sort. DNA doesn't replicate itself. Cells and, in real life, organisms copy their DNA using each strand of the double helix as the template for the synthesis of its partner. This process, of course, requires a whole series of physical and chemical conditions and reactions within the cell.

An important consequence of thinking of DNA as a "self-replicating" molecule, however, was that it sparked the imagination of a number of distinguished physicists and mathematicians, who, until then, had shown little interest in biological and biochemical systems and, indeed, perhaps a temperamental aversion to their inherent messiness. At the end of World War II, after two atomic bombs had been dropped on two Japanese cities, many physicists had become disillusioned with physics (harbinger of death) and were only too glad to turn their attention to biology (harbinger of life). Following the lead of the German exile and Nobel Prize physicist Erwin Schroedinger who, in his little book *What Is Life?* (Schroedinger 1944), had referred to the gene as a code and hailed it as the secret of life, they got excited about DNA. Familiar with wartime uses of cybernetics and code-breaking, they decided to try to crack the "genetic code" by devising formal solutions for the way different sequences of A, G, C, and T could get translated into the sequences of amino acids that constitute different proteins.

It is important to pay attention to the differences in the conceptual and physical tools the scientists attacking these questions used. As the messy biochemical work of grinding up tissues and isolating their cells and molecules yielded first place to the skills of code-breaking, centrifuges, spattered lab coats, and dirty glassware were replaced by paper, pencil, and soon by computers. In the process, different sequences of A, G, T, and C molecules became a "code" and the biological and chemical complexities of living organisms were reduced to abstractions about how to translate the linear "code" of DNA into the linear array of the amino acids that make up proteins.

In the process, what was conceptually pushed aside is the fact that this "trans-lation" ordinarily happens inside dividing and metabolizing cells of organisms, which live in complicated relationships with their environments. The complexities of such biological and social realities got erased as scientific interest focused on computations

and codes rather than on the interrelationships of gooey cells and A molecules, and, indeed, of the organisms and social structures among which life gets played out. And, although in the end the messy biochemists were the first to work out correspondences between the base sequences in DNA and the composition of proteins (for which they duly got their Nobel Prizes), much of the intellectual drama went with the more theoretical aspects of “breaking the code."

Before moving on, it is important to remember that, by itself, DNA is an inert, sticky glop. It takes organisms or, at least, the enzyme systems extracted from them, along with other essential molecules, to perform the synthetic processes within which DNA specifies either the composition of its own copies or the composition of proteins. As soon as we think of DNA as part of the living cells of living organisms, we realize that even a relatively simple trait, such as eye color, cannot possibly be “caused" by a single gene. Just to synthesize the pigments that color the iris of our eyes involves the participation of several proteins, the composition of each of which is specified by a different DNA sequence (or “gene"). Further proteins are required to knit the base sequences of these genes together and these proteins require further genes for their synthesis, and so on. Up to this point, we have not even begun to consider how the pigment gets deposited in the proper location in the iris or how our eyes, including the iris, get formed during embryonic development.

We are dealing with a situation in which even the “simplest” inherited trait about which we speak as though it were transmitted by a single gene, such as sickle cell disease or phenylketonuria (PKU), involves the participation of many proteins, and therefore of many “genes” (which is to say DNA sequences). The synthesis of these genes, in turn, requires further proteins, and so on and on. The usual shorthand of "the gene for" must not be taken literally. And yet, this way of thinking about genes has turned DNA into the “master molecule,” while proteins are said to fulfill “housekeeping” functions. (And one need not be a raving postmodernist to detect class, race, and gender biases in this way of describing the molecular relationships.)

Another level of complexity in the way DNA functions has to do with the fact that a “gene” - the piece of DNA that gets translated into a particular protein - often does not exist as a continuous base sequence on the chromosome. Presumably because of our long evolutionary history, a base sequence that specifies the composition of a given protein may be interrupted by sequences that were, until recently, thought to be meaningless gibberish. As a shorthand, molecular biologists sometimes call the coding (or “expressed") sequences - those that get translated into protein - exons, and the presumably meaningless sequences, introns. But, so far, no one understands how cells know to cobble appropriate exons together and to splice out the gibberish so as to produce the final sequence (or “message”) that specifies the composition of a particular protein. To make things even more complicated, exons often overlap or different parts of a given base sequence may function in different genes. In addition, pieces of expressed coding sequence can be buried inside what are thought to be meaningless introns. Such kinds of complexities have led many molecular biologists to stop using words like “gene,” “exon," or “intron” and to speak only of coding or noncoding sequences.

These sorts of largely unanticipated complexities suggest that the base sequence

of the human genome, which President Clinton hailed as “the language in which God created life" when it was announced with great fanfare in June 2000, is a very complex tongue, indeed. Of the string of bases that constitute the human genome, only some 3 percent are thought to be involved in specifying the composition of proteins. These 3 percent are by no means consecutive, and combinations of them can switch around or produce redundancies. Some of them, indeed, appear to get spliced into hundreds or even thousands of different “genes.” How the remaining 97 percent of bases function, or whether they even have a function, is as yet unknown.

It will take a long time to identify all the coding sequences and figure out how they combine to specify the composition of the many proteins, which function in the human body. It will also be no small task to understand how the relevant metabolic systems “decide” when, where, and at what rates different proteins are to be synthesized. The fact that the human genome turns out to harbor only about a third as many coding sequences as scientists had expected will make it all the more difficult to understand how coding sequences get cobbled together so as to perform all the functions attributed to them and to what extent they actually do so. The fact that the composition and number of the coding sequences of humans are quite similar to those found in mice (and even yeasts), despite the rather significant differences between us, will make it no easier to figure these relationships out.

In a sense, spelling out the sequences of A's, G’s, T's, and C's that constitute the human genome doesn’t put us conceptually that far ahead of where we were at the beginning of the twentieth century when biologists first decided that chromosomes and their genes play a fundamental role in the way cells and organisms are replicated, but had no idea how that might happen. At present, the translation of DNA into proteins seems straightforward only as long as we ignore the dynamic changes in which DNA, proteins, and our other body constituents participate from one moment to the next and in different locations in our bodies. There is no way even to imagine the extent and the ways DNA participates in the transformations our cells and bodies undergo in the course of our lives.

Unfortunately, these are not just interesting scientific or philosophical puzzles, because the contrast between the actual complexities and the conceptual simplifications scientists use when they try to explain them in terms of the dance of DNA has created a dangerous situation. Biotechnology - the industry of “genetic engineering" - is built on the pretense that scientists not only understand, but can anticipate and direct, the functions of the DNA sequences they isolate from organisms or manufacture in the laboratory. The industry cheerfully promises that it can foresee the potential effects of transferring specific DNA sequences, wherever and however obtained, into bacteria, plants, or animals, including humans, and thus improve targeted characteristics.

In reality, such operations can have three possible outcomes: (1) in the inhospitable environment of the cells of the host species, inserted DNA sequences do not succeed in specifying the intended proteins so that nothing new happens; (2) the inserted sequence mediates the synthesis of the desired protein product in the right amounts and at the right time and location; and (3) unpredicted and unintended consequences follow because the inserted DNA gets spliced into the wrong place in the genome of the host organism and disrupts or adversely alters one or more of its vital functions. The first alternative wastes

time and money, the second is the hope, but the third spells danger. Yet, which of them happens cannot be predicted a priori, or from one genetic manipulation to another, since the conditions within and around
the host organisms are likely to change over time.

Clearly, the model underlying the promise of genetic engineering is overly simplistic. But, what makes the situation even more problematic is that DNA sequences, once isolated or synthesized, as well as the cells, organs, or organisms into which they are inserted can be patented and thereby become forms of intellectual property. The science and business of genetic engineering have become one, and efforts at basic understanding compete with the pursuit of profits. The usual professional rivalries are enhanced by major financial rivalries, and the complete interlinking of government, universities, and industry leaves hardly any disinterested scientists, devoid of conflicts of interest, who can be trusted to evaluate and critique proposed scientific models or their practical implementation without raising suspicions of pursuing financial interests. As the biotechnology industry expands its reach, the health hazards and environmental pollution it produces are added to those chemistry and physics have bequeathed us during the twentieth century.

In this essay, I have tried to hint at the complex dialectical relationships among the material, ideological, social, political, and economic dimensions and implications of the supposedly scientific gene concept. The gene, in fact, is a prime example of what Niels Bohr referred to as complementarity.

Bohr initially formulated this concept to denote the fact that electromagnetic radiation can be pictured as both waves and particles. Contradictory as those representations seem, it is not one or the other, but both at all times. Which of them constitutes the appropriate description simply depends on what instruments are used to detect it. Similarly, genes are DNA molecules, but they also are symbols of health and disease, of hopes and fears for the future, of scientific fame and dishonor, of business fortunes and failures, and no doubt much more. To ignore any of these aspects leaves the gene concept incomplete. A central icon of our time, the gene is simultaneously a material object and an ideology, full of political, economic, spiritual, individual, and societal content.

References

Baldwin, E. 1948. *Dynamic aspects of biochemistry*. New York: Macmillan.
Cannon, W. B. 1932. *The wisdom of the body*. New York: W. W. Norton.
Loeb, J. 1916. *The organism as a whole.* New York: G.P. Putnam's Sons.
Maddox, B. 2002. *Rosalind Franklin*. New York: HarperCollins.
Schroedinger, E. 1945. *What is life?* New York: Macmillan.
Watson, J.D. 1968. *The double helix.* New York: Athenaeum.
Watson, J.D., and F.H.C. Crick, 1953. A structure for deoxyribonucleic acid. *Nature* 171: 737-8.