COMMENT

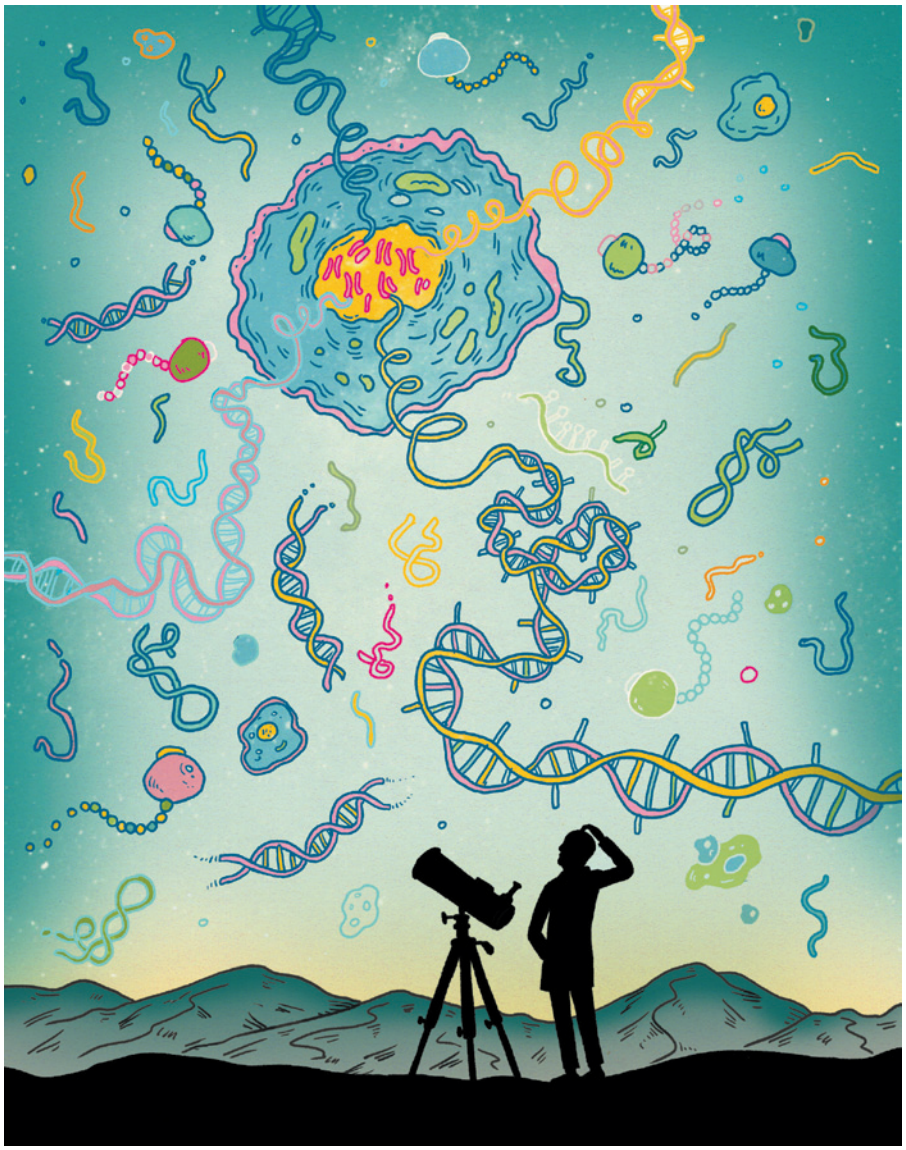
RECOVERY Lab devastated by Hurricane Sandy picking up the pieces six months on **p.421**

NEUROSCIENCE Spring Books special opens with Douglas Hofstadter's latest **p.424**

CULTURE Steve Jones' new book explores the science invoked by the Bible **p.432**

DNA Newfound letter sheds light on nominations behind double-helix Nobel **p.434**

ANDREW RAE



Celebrate the unknowns

On the 60th anniversary of the double helix, we should admit that we don't fully understand how evolution works at the molecular level, suggests **Philip Ball**.

This week's diamond jubilee of the discovery of DNA's molecular structure rightly celebrates how Francis Crick, James Watson and their collaborators launched the 'genomic age' by revealing how hereditary information is encoded in the double helix. Yet the conventional narrative — in which their 1953 *Nature* paper led inexorably to the Human Genome Project and the dawn of personalized medicine — is as misleading as the popular narrative of gene function itself, in which the DNA sequence is translated into proteins and ultimately into an organism's observable characteristics, or phenotype.

Sixty years on, the very definition of 'gene' is hotly debated. We do not know what most of our DNA does, nor how, or to what extent it governs traits. In other words, we do not fully understand how evolution works at the molecular level.

That sounds to me like an extraordinarily exciting state of affairs, comparable perhaps to the disruptive discovery in cosmology in 1998 that the expansion of the Universe is accelerating rather than decelerating, as astronomers had believed since the late 1920s. Yet, while specialists debate what the latest findings mean, the rhetoric of popular discussions of DNA, genomics and evolution remains largely unchanged, and the public continues to be fed assurances that DNA is as solipsistic a blueprint as ever.

The more complex picture now emerging raises difficult questions that this outsider knows he can barely discern. But I can tell that the usual tidy tale of how 'DNA makes RNA makes protein' is sanitized to the point of distortion. Instead of occasional, muted confessions from genomics boosters and popularizers of evolution that the story has turned out to be a little more complex, there should be a bolder admission — indeed a celebration — of the known unknowns.

DNA DISPUTE

A student referring to textbook discussions of genetics and evolution could be forgiven for thinking that the 'central dogma' devised by Crick and others in the 1960s — in which information flows in a linear, traceable fashion from DNA sequence to messenger RNA to protein, to manifest finally as phenotype — remains the solid foundation of the genomic revolution. In fact, it is beginning to look more like a casualty of it.

Although it remains beyond serious doubt that Darwinian natural selection drives much, perhaps most, evolutionary change, it is often unclear at which phenotypic level selection operates, and particularly how it plays out at the molecular level.

Take the Encyclopedia of DNA Elements (ENCODE) project, a public research consortium launched by the US National Human Genome Research Institute in Bethesda, Maryland. Starting in 2003, ENCODE researchers set out to map which parts of human chromosomes are transcribed, how transcription is regulated and how the process is affected by the way the DNA is packaged in the cell nucleus. Last year, the group revealed[1] that there is much more to genome function than is encompassed in the roughly 1% of our DNA that contains some 20,000 protein-coding genes — challenging the old idea that much of the genome is junk. At least 80% of the genome is transcribed into RNA.

Some geneticists and evolutionary biologists say that all this extra transcription may simply be noise, irrelevant to function and evolution[2]. But, drawing on the fact that regulatory roles have been pinned to some of the non-coding RNA transcripts discovered in pilot projects, the ENCODE team argues that at least some of this transcription could provide a reservoir of molecules with regulatory functions — in other words, a pool of potentially 'useful' variation. ENCODE researchers even propose, to the consternation of some, that the transcript should be considered the basic unit of inheritance, with 'gene' denoting not a piece of DNA but a higher-order concept pertaining to all the transcripts that contribute to a given phenotypic trait[3].

According to evolutionary biologist Patrick Phillips at the University of Oregon in Eugene, projects such as ENCODE are showing scientists that they don't really understand how genotypes map to phenotypes, or how exactly evolutionary forces shape any given genome.

COMPLEX CODE

The ENCODE findings join several other discoveries in unsettling old assumptions. For example, epigenetic molecular alterations to DNA, such as the addition of a methyl group, can affect the activity of genes without altering their nucleotide sequences. Many of these regulatory chemical markers are inherited, including some that govern susceptibility to diabetes and cardiovascular disease[4]. Genes can also be regulated by the spatial organization of the chromosomes, in turn affected by epigenetic markers. Although such effects have long been known, their prevalence may be much greater than previously thought[5].

Another source of ambiguity in the genotype–phenotype relationship comes from the way in which many genes operate in complex networks. For example, many differently structured gene networks might result in the same trait or phenotype[6]. Also, new phenotypes that are viable and potentially superior may be more likely to emerge through tweaks to regulatory networks than through more risky alterations to protein-coding sequences[7]. In a sense this is still natural selection pulling out the best from a bunch of random mutations, but not at the level of the DNA sequence itself.

One consequence of this complex genotype–phenotype relationship is that it may impose constraints on natural selection. If the same phenotypes can result from many similarly structured gene networks, it might take a long time for a 'fitter' phenotype to arise[8]. Alternatively, mutations may accumulate, free from selective 'weeding', thanks to the robustness of networks in maintaining a particular phenotype. Such hidden variation might be unmasked by some new environmental stress, enabling fresh adaptations to emerge[9]. These sorts of constraints and opportunities are poorly understood; evolutionary theory does not help biologists to predict what kinds of genetic network they should expect to see in any one context.

"Simplistic portrayals of evolution encourage equally simplistic demolitions."

Researchers are also still not agreed on whether natural selection is the dominant driver of genetic change at the molecular level. Evolutionary geneticist Michael Lynch of Indiana University Bloomington has shown through modelling that random genetic drift can play a major part in the evolution of genomic features, for example the scattering of non-coding sections, called introns, through protein-coding sequences. He has also shown that rather than enhancing fitness, natural selection can generate a redundant accumulation of molecular 'defences', such as systems that detect folding problems in proteins[10]. At best, this is burdensome. At worst, it can be catastrophic.

In short, the current picture of how and where evolution operates, and how this shapes genomes, is something of a mess. That should not be a criticism, but rather a vote of confidence in the healthy, dynamic state of molecular and evolutionary biology.

A PROBLEM SHARED

Barely a whisper of this vibrant debate reaches the public. Take evolutionary biologist Richard Dawkins' description in *Prospect* magazine last year of the gene as a replicator with "its own unique status as a unit of Darwinian selection". It conjures up the decades-old picture of a little, autonomous stretch of DNA intent on getting itself copied, with no hint that selection operates at all levels of the biological hierarchy, including at the supraorganismal level[2], or that the very idea of 'gene' has become problematic.

Why this apparent reluctance to acknowledge the complexity? One roadblock may be sentimentality. Biology is so complicated that it may be deeply painful for some to relinquish the promise of an elegant core mechanism. In cosmology, a single, shattering fact (the Universe's accelerating expansion) cleanly rewrote the narrative. But in molecular evolution, old arguments, for instance about the importance of natural selection and random drift in driving genetic change, are now colliding with questions about non-coding RNA, epigenetics and genomic network theory. It is not yet clear which new story to tell.

Then there is the discomfort of all this uncertainty following the rhetoric surrounding the Human Genome Project, which seemed to promise, among other things, 'the instructions to make a human'. It is one thing to revise our ideas about the cosmos, another to admit that we are not as close to understanding ourselves as we thought.

There may also be anxiety that admitting any uncertainty about the mechanisms of evolution will be exploited by those who seek to undermine it. Certainly, popular accounts of epigenetics and the ENCODE results have been much more coy about the evolutionary implications than the developmental ones. But we are grown-up enough to be told about the doubts, debates and discussions that are leaving the putative 'age of the genome' with more questions than answers. Tidying up the story bowdlerizes the science and creates straw men for its detractors. Simplistic portrayals of evolution encourage equally simplistic demolitions.

When the structure of DNA was first deduced, it seemed to supply the final part of a beautiful puzzle, the solution for which began with Charles Darwin and Gregor Mendel. The simplicity of that picture has proved too alluring. For the jubilee, we should do DNA a favour and lift some of the awesome responsibility for life's complexity from its shoulders. ■

Philip Ball *is a freelance science writer based in London.*
e-mail: p.ball@btinternet.com

1. The ENCODE Project Consortium *Nature* **489,** 57–74 (2012).
2. Doolittle, W. F. *Proc. Natl Acad. Sci. USA* **110,** 5294–5300 (2013).
3. Djebali, S. *et al. Nature* **489,** 101–108 (2012).
4. Jablonka, E. & Raz, G. *Q. Rev. Biol.* **84,** 131–176 (2009).
5. Mattick, J. S. *Proc. Natl Acad. Sci. USA* **109,** 16400–16401 (2012).
6. Wagner, A. *Trends Genet.* **27,** 397–410 (2011).
7. Mattick, J. S. *FEBS Lett.* **585,** 1600–1616 (2011).
8. Wagner, A. *Trends Ecol. Evol.* **26,** 577–584 (2011).
9. Jarosz, D. F. & Lindquist, S. *Science* **330,** 1820–1824 (2010).
10. Lynch, M. *Proc. Natl Acad. Sci. USA* **109,** 18851–18856 (2012).

Cell

1953: When Genes Became "Information"

Matthew Cobb[1,*]
[1]Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, UK
*Correspondence: cobb@manchester.ac.uk
http://dx.doi.org/10.1016/j.cell.2013.04.012

In 1953, Watson and Crick not only described the double-helix structure of DNA, but also embraced the idea that genes contained a code that expresses information and thereby changed our view of life. This article traces how these ideas entered biological thinking and highlights the connections between different branches of science at the time, exploring the power of metaphor in science.

Introduction

Sixty years ago, James Watson and Francis Crick described the double-helix structure of DNA. The double helix famously led them to state that "it has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material." However, replication is merely one aspect of the hereditary substance, one that is shared by some forms of nonliving matter, like crystals. The key thing about the DNA molecule was revealed in their next paper, which appeared in *Nature* 6 weeks later (Watson and Crick, 1953). The conceptual breakthrough they announced in this second paper changed humanity's vision of life: "it therefore seems likely that the precise sequence of the bases is the code which carries the genetical information." Sixty years on and shorn of the opening conditional phrase, something like these words is uttered in biology classes around the world every single day.

The story of how Watson, Crick, and so many others came to hunt for the structure of DNA is well known (Watson, 2012). What is far less well understood is how Watson and Crick came to use those key ideas that look so obvious to us now—"code" and "information." In 1940—when Oswald Avery began his final search for the "transforming agent" in bacteria, which, in 1944, would culminate in the discovery that the hereditary material is DNA—neither of these ideas existed as biological concepts. The way these fundamental ideas entered biology in the 1940s reveals surprising interconnections between different aspects of science at the time.

Schrödinger's Code

In February 1943, the Nobel-Prize-winning physicist Erwin Schrödinger gave a series of public lectures in Dublin that were later collected under the title *What Is Life?* In these lectures, Schrödinger looked at recent findings in biology, including the nature of mutations and the size of genes (Schrödinger, 1944). The exact role of Schrödinger's book in the development of molecular biology has been the subject of argument among participants and historians (e.g., Kay, 2000; Pauling, 1987; Yoxen, 1979). Whatever the case, the book did inspire many of those who came to focus their lives on the structure of DNA and of genes—James Watson, Francis Crick, Maurice Wilkins, and Seymour Benzer, among others. Above all, Schrödinger was the first person in the 20th century to explicitly suggest that genes contained what he called a "code-script" that determined "the entire pattern of the individual's future development and of its functioning in the mature state."

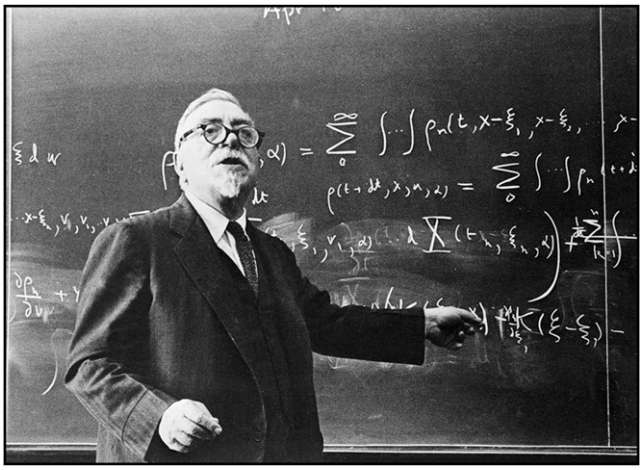
Schrödinger's idea looks so prescient that it is tempting to conclude that it must have directly influenced Watson and Crick's thinking a decade later. But neither Watson and Crick, nor indeed many people, appear to have been struck by Schrödinger's code-script idea. The term was noted in *Nature*, but simply as a synonym for "a genome," and in the *New York Times*, but merely as something "that gives orders which are carried out."

In an early attempt to investigate the idea that genes contain a "code," Kurt Stern developed a model of helical nucleoprotein molecules in which the nucleic acid chains were modulated in combination with polypeptide chains, similar to "the modulations impressed on a smooth surface by the stylus of a sound recorder" (Stern, 1947). At around the same

"Spirals Time—Time Spirals" by Charles Jencks is a large sculpture of a DNA double helix, and it was donated by Jim Watson to Cold Spring Harbor Laboratory in 2000. Image courtesy of Chua/CSHL.

time, Erwin Chargaff not only identified the proportions of the different bases in samples of nucleic acid, but also began to consider the nature of the code. At the 1947 Cold Spring Harbor Symposium, Chargaff suggested that "differences in the proportions or in the sequence of the several nucleotides forming the nucleic acid chain also could be responsible for specific effects" (Chargaff, 1947). By 1950, Chargaff was explicitly arguing that this was the case: "We must realize that minute changes in the nuclear acid, e.g., the disappearance of one guanine molecule out of a hundred, could produce far-reaching changes in the geometry of the conjugated nucleoprotein, and it is not impossible that rearrangements of this type are among the causes of the occurrence of mutations" (Chargaff, 1950).

On closer inspection, Schrödinger's use of "code" was not at all like that of Watson and Crick in 1953—he did not think that there was a correspondence between each part of the gene and precise biochemical reactions, which is what a code implies (Kay, 2000). Although Beadle and Tatum had recently shown that, in *Neurospora*, different mutations affected specific enzymes, Schrödinger was apparently unaware of their work. So although Schrödinger used the term code and explicitly suggested that the hereditary molecule (which, like virtually everyone else in 1943, he assumed was a protein) could contain massive amounts of variability that could act as the basis of that code, he did not address the issue of what exactly the code-script contained. One reason for this difference between Schrödinger's conception of code and that used by Watson and Crick (1953) was that, during the 10 years of tumultuous discovery and conceptual exploration that separated Schrödinger's lectures from the second 1953 *Nature* paper, a key metaphor, completely absent from Schrödinger's thinking, entered into science: *information*.



Norbert Wiener (1894–1964), the founder of "Cybernetics," was a brilliant mathematician who realized that most organic systems are composed of feedback loops and also contributed to the recognition of information as a decisive feature of all systems.

Enter Information

"Information" became a key scientific concept through the work of the US National Research and Development Committee, which was set up by President Roosevelt to fund scientific research into military problems. It was initially divided into four sections, including one that would later become the Manhattan Project. Section D-2 studied "fire control"—how to ensure accurate artillery fire, in particular antiaircraft fire, by the integration of information from radar, visual tracking, and range finding. The director of D-2 was Warren Weaver, the mathematical physicist who ran the Rockefeller Institute and had coined the term "molecular biology."

Among the men Weaver worked with was Claude Shannon, a mathematician from Bell Labs who had recently obtained his PhD in the "algebra of theoretical genetics" and was interested in both fire control and codes. Shannon was interested in what he called "discrete information" and the way that information often contained redundancy. Another significant figure in Weaver's team was the mercurial Norbert Wiener, a mathematician with an interest in control systems. Wiener developed statistical approaches to understand how the antiaircraft system should respond to evasive action by the aircraft, focusing on the control interface between man and machine. Both Shannon and Wiener shared a central interest in the role of feedback, control, and the importance of information.

In 1944, Wiener teamed up with computer pioneer John von Neumann and organized a series of meetings in which engineers who studied control and communication could discuss with scientists studying biomedical questions. Funded by the Rockefeller Institute and the Macy Foundation, these informal meetings marked the beginning of a sea change in the way scientists viewed behavior and physiology. The concepts of control, feedback, and information began to seep into the scientific vocabulary, words that would soon enable new hypotheses to be generated and new experiments to be devised.

In 1948, the shift in thinking that had been developing during the war years finally exploded into the public domain in three phases. First, Wiener published *Cybernetics: Or Control and Communication in the Animal and the Machine*, in which he proclaimed "the present time is the age of communication and control." Messages, codes, and information were at the heart of Wiener's vision, which was acclaimed by academics and by newspapers across the globe. Shortly afterward, Shannon and Weaver published *The Mathematical Theory of Information*, in which they developed a general mathematical framework that could apply to any system, organic or inorganic, living or electrical. Finally, at the 1948 Hixon Symposium held in California, which was attended by biologists such as Delbrück and Beadle, von Neumannn envisaged a biological equivalent of a computing machine able to reproduce itself with a set of instructions corresponding to a gene; in von Neumann's striking view, a gene was seen as an "information tape" that could program the organism—like the "universal Turing machine" described in 1936 by Alan Turing.

These approaches began to influence thinking about what exactly genes contain. In his popular 1950 book about cybernetics, *The Human Use of Human*

Beings, Wiener argued that genes constituted a kind of "memory" that was "transmitted" (yet another electronic term). Information was the essence of life, Wiener was now arguing, and the hereditary material (which he assumed to be proteins) was responsible for transmitting that information. In 1950, Hans Kalmus explicitly developed this idea in an article entitled "A Cybernetical Aspect of Genetics," in which he described the gene as a "message" of a "chemical nature" (Kalmus, 1950). This was met with a resounding silence—the first citation of the article was in 1962 by Kalmus himself.

The power of the information or message metaphor now looks obvious, but at the time its impact was hindered by the fact that it did not really explain anything. Instead, it merely emphasized the fundamental problem that so many minds were concentrating on—the physical structure of the hereditary material. Stating that a gene contained—or was—information did not really help in understanding the key issue of the day, which was how genetic specificity (or information) was encoded in the DNA molecule. Furthermore, attempts to apply the more specific aspects of Shannon's information theory to biological problems proved difficult. In 1949, Henry Quastler estimated the "information content" of a human being at about 5×10^{25} bits (Kay, 1995). At the same time, in an unpublished sketch, Shannon suggested that the "genetic constitution of man" contained slightly less than 10^5 bits of information (Gleick, 2011). Neither of these estimates was based on anything more than guesswork.

An indication of what some of the leading molecular biologists thought of the fashion for cybernetics and information theory can be gleaned from a letter entitled "Terminology in Bacterial Genetics," which appeared in *Nature* the week before Watson and Crick published their April 1953 paper. The letter—which was a spoof—had been cooked up in September 1952 over a well-oiled lunch by Boris Ephrussi, Jim Watson, and others (Watson, 2001). They satirically suggested that various terms used in bacterial genetics should be replaced by "interbacterial information" and closed with a reference to "the possible future importance of cybernetics at the bacterial level" (Ephrussi et al., 1953). The editor did not get the joke (to be fair, it was not very funny), and he published the letter, leading historians to take the apparently brilliant insight equally seriously (e.g., Kay [1995] considered that the letter represented a "gestalt switch"; even if it was a joke, she was probably right).

Only 7 weeks separated the Ephrussi et al. (1953) letter from Watson and Crick's second 1953 paper, but the conceptual gulf was vast. Ephrussi and Watson's mickey taking was replaced by serious, revolutionary science. In the May 30, 1953 issue of *Nature*, Watson and Crick addressed "the genetical implications of the structure of deoxyribonucleic acid" and used the concept of information in a way that fully expressed the radical nature of their discovery—"the precise sequence of the bases is the code which carries the genetical information." It is not known where this powerful phrase came from—Watson recalls that the paper was written in a rush, in less than a week, in an attempt to develop aspects of their model that had been deliberately unstated in the first paper (Watson, 2001). Whatever the case, in the following years, the idea of information and of a code was at the heart of the key developments in genetics and molecular biology, as various attempts were made to crack what was now called the genetic code, beginning with physicist George Gamow's July 1953 letter to Watson and Crick (Watson, 2001) and culminating in 1961 with Marshall Nirenberg and Heinrich Matthaei's brilliant "poly-U" experiment.

What Happened Next

Although code and information became commonplace metaphors in biological thinking in the 1950s, they were—and are—merely vague ways of interpreting genetic phenomena rather than precise theoretical frameworks. In fact, Shannon's strict version of information turned out to be pretty much of a biological dead end, for the time being at least. On the other hand, information as a metaphor, with genetic information having an instructional nature, has survived and flourished. Cybernetics similarly failed to fulfill its boastful promises of integrating all levels of biology; however, its emphasis on control and feedback loops did prove extremely influential.

Jacques Monod, in particular, embraced the cybernetical approach—after the "PaJaMa" experiments showed that a single regulator gene controlled the activity of several structural genes through feedback loops, Monod explicitly cast his work in the language of cybernetics. In 1959, he even began writing a book called *Enzymatic Cybernetics* (Kay, 2000). Ephrussi and Watson's satirical wheeze that had fooled the editor of *Nature* 6 years earlier had become reality—cybernetics was being used to understand bacteria. However, as with information theory, it was the general framework, rather than the precise mathematical detail, that was being employed. Cybernetics became an analogy, a metaphor, a way of thinking about biological processes, rather than a new science.

The final step in the link between biology and the most recent developments in electronic technology took place in 1961 when Jacob and Monod summarized their view of gene function and protein synthesis. They used terms that are both utterly modern and harked back almost word for word to Schrödinger's view of the nature of the code-script: "the genome contains not only a series of blueprints, but a coordinated program of protein synthesis and the means of controlling its execution" (Jacob and Monod, 1961). The gene had entered the computer age. According to this metaphor, not only did the genetic code contain information, but this information had a special kind of meaning—it was a program, a set of instructions that enabled the cell to carry out a particular activity. However, it is easy to forget that this is a figure of speech rather than being literally true. A gene is like a program, but it is not a computer program and does not function according to the same rules. Similarly, the genetic code is not literally a code—it is a process that enables organisms to carry out particular functions by turning stored information into structures or actions using evolved rules.

Science proceeds primarily by evidence rather than by theory, and experimentation is generally the most powerful way of obtaining conclusive evidence. But, to interpret this evidence, we need theories and conceptual frameworks,

which in turn are made up of words, metaphors, and analogies that are often based on the most recent technological developments. This is powerfully shown by the fusion of Schrödinger's code-script and Shannon's information, which occurred in the lucid prose of Watson and Crick in their second article of 1953. As they embraced this radical framework and this new way of seeing, they transformed our vision of life.

ACKNOWLEDGMENTS

This article forms part of a work in progress on the history of the genetic code, to be published by Profile Books in 2014.

REFERENCES

Chargaff, E. (1947). Cold Spring Harb. Symp. Quant. Biol. *12*, 28–34.

Chargaff, E. (1950). Experientia *6*, 201–209.

Ephrussi, B., Leopold, U., Watson, J.D., and Weigle, J.J. (1953). Nature *171*, 701.

Gleick, J. (2011). The Information: A History, A Theory, A Flood (London: Fourth Estate).

Jacob, F., and Monod, J. (1961). J. Mol. Biol. *3*, 318–356.

Kalmus, H. (1950). J. Hered. *41*, 19–22.

Kay, L.E. (1995). Sci. Context *8*, 609–634.

Kay, L.E. (2000). Who Wrote the Book of Life? A History of the Genetic Code (Stanford: Stanford University Press).

Pauling, L. (1987). Schrödinger's contribution to chemistry and biology. In Schrödinger: Centenary Celebration of a Polymath, C.W. Kilmister, ed. (Cambridge: Cambridge University Press), pp. 225–233.

Schrödinger, E. (1944). What is Life? (Cambridge: Cambridge University Press).

Stern, K.G. (1947). Yale J. Biol. Med. *19*, 937–949.

Watson, J.D. (2001). Genes, Girls and Gamow (Oxford: Oxford University Press).

Watson, J.D. (2012). The Annotated and Illustrated Double Helix, A. Gann and J. Witkowski, eds. (London: Simon & Schuster).

Watson, J.D., and Crick, F.H.C. (1953). Nature *171*, 964–967.

Yoxen, E.J. (1979). Hist. Sci. *17*, 17–52.