Now What?

John J. Hopfield

October 2018

Now What?

John J. Hopfield, Princeton University, Princeton, NJ 08544, USA.

October 2018

John J. Hopfield Now What? (Article in PDF format)

My first full-time permanent employment was at the Bell Telephone Laboratories in Murray Hill NJ, where the transistor had been invented 10 years earlier. Its six-person theoretical physics group had hired me as a Member of Technical Staff. My physics thesis written and approved, and all other Ph.D. requirements completed at Cornell University, I reported to work early in March 1958. A first half-day of employment was spent on administrative details. I joined a couple of theorists in the cafeteria for lunch, and then headed for my new office. Unpacking a few books and journals took an hour. A visit to the 5th floor stockroom produced some lined tablets, pencils, and a hand-held pencil sharpener. I sharpened several pencils.

Now What?

That stark question has plagued me many times in my research and teaching career. Every scientist, every scholar, every writer, every artist. . . faces this vexing issue of what to work on each day. Most respond by working today to extend a little yesterday's line of thought, or line of measurement, or plot development or . . . Most, if forced to ask "Now What" at a more fundamental level, or for a long-term commitment of effort, endeavor to not much change the sandbox in which they have been happily playing.

My 2019 Franklin Medal in Physics is:

> *For applying concepts of theoretical physics to provide new insights on important biological questions in a variety of areas, including genetics and neuroscience, with significant impact on machine learning, an area of computer science. (Franklin Institute 2018)*

Those came about because I have often responded to the "Now What" question by a major change of direction.

The more scientific citation for the Franklin award indicates that two of my research papers were central to my nomination. This essay describes the history of how I happened upon the research problems these papers address or clarify. My answer to "Now What" is "here is a research problem which is unusual, perhaps significant, novel, that I can pose and probably solve because of my background in physics". The situation would not be readily identified as a problem at all by those whose background seems much more relevant than my own.

Choosing problems is the primary determinant of what one accomplishes in science. I have generally had a relatively short attention span in science problems (note the fine line between the polymath and the dilettante, where I am often offside). Thus I have always been on the lookout for more interesting questions either as my present ones get worked out, or as they get classified by me as intractable, given my particular talents.

What is physics?

To me–growing up with a father and mother both of whom were physicists—physics was not subject matter. The atom, the troposphere, the nucleus, a piece of glass, the washing machine, my bicycle, the phonograph, a magnet–these were all incidentally the subject matter. The central idea was that the world is understandable, that you should be able to take anything apart, understand the relationships between its constituents, do experiments, and on that basis be able to develop a quantitative understanding of its behavior. Physics was a point of view that the world around us is, with effort, ingenuity, and adequate resources, understandable in a predictive and reasonably quantitative fashion. Being a physicist is a dedication to a quest for this kind of understanding.

Education, broadly construed

I grew up taking things apart, seeing how they worked, repairing bicycles, exploring chemistry in the kitchen (or better, out of sight in the cellar), building flyable model airplanes, crystal sets and simple radios, playing with batteries and coils of wire, and learning to think with my hands and manipulate real objects. One of my earliest memories is of a small screwdriver that was kept in a drawer of the treadle–operated sewing machine my mother used. It was for minor sewing machine adjustments, but I was allowed to use it on anything in the house—as long as I put it back in the drawer. And if I occasionally could not reassemble the object I had attacked, my father would patiently do so in the evening. My early concept of what it would be like to be a physicist was a somewhat mystical idea of carrying on such playful explorations at a more abstract level.

In my father's generation, physics certainly did not involve being well paid. My father went to Berlin on a Guggenheim Fellowship in 1929, planning to return to the US in 1930 to take an academic job. The Wall Street crash of 1929 intervened, eliminating all university hiring. Instead, he took a sequence of temporary positions, including setting up the physics exhibit at the 1933 World's Fair in Chicago, where I was born. He ultimately took a position with the Libby Owens Ford glass company, where he worked on tempered glass and thermopane, both of which turned out to be materials science problems. Heedless of the financial consequences of being a physicist, I only briefly considered an alternative career. My high school chemistry teacher had been superb, while my physics teacher (in hindsight) lacked an understanding of even the rudiments of mechanics and electricity. Thus influenced, when listing potential majors on college applications I wrote 'physics or chemistry'.

Arriving at Swarthmore I was assigned William Elmore as my advisor. Bill knew my father and my background. I still remember going into his office for the first time, when he picked up the card describing my interests and the fields of study I was considering, took out a pen, and crossed out two words, remarking 'I don't believe we need to consider chemistry'.

Ultimately one must choose something particular on which to do research. So late in my Swarthmore career I began looking at graduate schools in physics with an eye to what I might specialize in. My upbringing had particularly focused my interest on the physics of the world around me, not the physics of the nucleus or the cosmos, but rather the physics of the daily world and its technologies. I eventually went to graduate school at Cornell, rather than Princeton, because in 1954 Cornell seemed to have a segment of the department interested in a field called solid-state physics. It was defined by a few courses with solid-state in their titles, a solid-state seminar, two faculty doing theoretical research in this area, and about four experimental thrusts. Experimental problems being worked on included low temperature thermal conductivity, color centers in the alkali halides, the UV spectroscopy of insulators, and X-ray absorption. The one person doing research on super fluidity of 4He was certainly not within the scope of solid-state physics as defined at the time. Donald Holcomb, recently arrived from the University of Illinois had a Varian NMR rig, and was on the fringe. The other established interests of the Department were nuclear and particle physics, x-rays, and cosmology/astrophysics.

In the middle of my second year at Cornell I approached theoretical physicist Albert W. Overhauser to ask if he would supervise my thesis and help me find a thesis problem. Through courses and doing problems I was rapidly acquiring the tools of theoretical physics, but I had no idea how to find an appropriate research problem. Finding a good problem was not (and is not even today) a subject discussed in classes or seminars.

Happily, Al had a long list of interesting puzzles. These were often in the form of a paradox 'elementary theoretical analysis of a particular phenomenon in solid-state physics gives result A, while experiments give an entirely different result B'. For example, the opening two sentences of the most cited paper with Overhauser as an author, are

> The simple classical theories of the dielectric constants and compressibility of ionic crystals lead to two relations among the experimental quantities from which arbitrary parameters have been eliminated, [called] the Szigeti relations. Neither is satisfied by the data, indicating the inadequacy of these simple theories. (B. G. Dick and A. W. Overhauser, 1957)

The paper, summarizing the Ph.D. thesis of Overhauser's second student Gayle Dick, describes the physics of a much better model that eliminates the discrepancy between theory and experimental results.

In most of paradoxes in his list Overhauser himself had no idea of what was producing the conflict between a commonsense theoretical approach to an answer and the reality coming from experiments. I picked one having to do with the radiative lifetime of an exciton in a crystal, where the conflict was within theory itself. Naïve theory yielded either zero or infinity depending on how it was applied, neither of which seemed to make sense. It became my problem, and Overhauser never worked on it at all. He was enormously supportive as a listener and critic when I went to see him, but finding direction and resolving technical theoretical issues were entirely my problem. The great gift he gave me was ownership of an interesting question, and total responsibility for research and progress. One day he simply told me that I had better start writing up my understanding, and that was the end of my Ph.D. research. The polariton, a new solid state physics 'particle', was invented to resolve the paradoxical situation. The single paper (and single author) paper written from my thesis (Hopfield, 1958) is still highly cited thanks to lasers, the polariton condensate, and modern photonics. Thank you, Al. I have done my best to repay you through similarly nurturing another generation of independent students.

First real employment

With a theoretical solid-state thesis in hand, I went into the job market. Academia or industry? The answer was clear from the Solid State Seminar speakers I had heard. AT&T's Bell Labs (Murray Hill) and General Electric (Schenectady) had broader, more vigorous research programs in solid state physics than any university did. I came home from visits to these laboratories positively euphoric about both the field as it was pursued in industry and the relatively unfettered research climate of both laboratories. I would ultimately go to Bell Labs chiefly because of its laboratory administrative structure, which had been reworked so that there was a small theoretical physics group not directly assigned to a sub-field such as magnetism or semiconductors. By contrast, GE had a couple theorists in each of its solid-state physics topical subgroups. And of course, joining Bell Labs in solid-state theory was a little like joining the Yankees as a pitcher in that era.

The theorists were all working on problems similar in kind to those that had motivated Overhauser. P. W. Anderson had just written his "Absence of diffusion in certain random lattices" paper. It would form the basis of his Nobel Prize and was motivated by an effort to explain some paradoxical electron spin relaxation results in Feher's experiments in doped silicon. Melvin Lax was trying to formulate the noise problem in semiconductor diodes in such a way that it did not violate the second law of thermodynamics. (It is very difficult to write a theory in which the diode rectification characteristic results in zero average charge on a capacitor in a simple diode-resistor circuit.) Conyers Herring was inventing "phonon drag" to explain an anomalously huge thermoelectric power in doped semiconductors. Gregory Wannier was working on 'Stark ladders' and their possible observation at high electric fields. What strikes me, looking back now at the whole enterprise, was that the subjects of study were generic problems of solid-state physics. The theoretical efforts were often based on detailed experiments, but they were not specifically driven by the materials science and device needs of AT&T. The bulk of the Bell Labs effort was of course on problems of that latter kind, but the theoretical group (sub-department 1111) was different.

Herring read all the solid-state literature, and even himself did some of the translations from the Soviet literature. He ran a monthly Journal Club, selecting the most interesting of his readings for presentations that he assigned to relevant Labs scientists. Few declined to do so. These highly argumentative meetings were

marvelous for my extended education and as an introduction to the extremely diverse society that comprised Bell Labs.

Conyers was also the head of the theory group, and saw me struggling to identify a next problem to work on. He suggested visiting with experimentalists to get my own sense of where interesting puzzles lay, and I still remember many of those visits. Arthur Schawlow explained the 1958 Townes-Schawlow theoretical paper on the possibility an 'optical maser'. He then took from his desk drawer a pink rod about three inches long, and described the wonderful intricacies of the spectroscopy of $Cr3+$ impurities which made $Al2O3$ into a ruby. He finished his discourse by remarking that unfortunately the R1 and R2 spectral lines were totally unsuitable to making an optical maser. (Two years later, after Maiman had shown lasing action in a flash-lamp pumped ruby, the crystal was put to use in the first Bell Labs built laser. The significance of making a pulsed laser had not occurred to Schawlow). Bill (Willard) Boyle was measuring the low-temperature magneto-thermal properties of the semi-metal bismuth. George Feher had mapped out the electronic wave function of phosphorous donors in silicon by the ENDOR technique that he had recently developed. Bernd Matthias was telling all who would listen that superconductivity could not really be BCS-like because of the absence on an isotope effect in some materials. Robert J. Collins introduced me to "edge emission" luminescence in CdS which provided a topic for my second paper.

A visit with Jim Lander's sub department, which was totally devoted to ZnO, introduced me to David G. Thomas and our subsequent collaboration was to provide me with the necessary puzzles for the next few years. The mystique of the golden days of Bell Labs and the wonderful foresight of its administrators and scientists leads me to relate the back-story of this group. Prior to 1960 vacuum tubes were the basic amplifiers in all telephone electronics. Vacuum tubes required electron emission from hot cathodes, but for long life and low power consumption the temperature needed to be as low as possible. BaO coated cathodes worked well in this regard, so Bell Labs had developed a group investigating BaO.

By the mid-1950's it was apparent that the day of vacuum tubes was coming to an end. In spite of this fact, the first transatlantic cable carrying telephone conversations (1956) had vacuum tube amplifiers every 43 miles. (The cable was taken out of service because of technical obsolescence after 22 years of use, with all the vacuum tubes still working!) With BaO no longer a material of future technological interest to AT&T, and because the group had an experience base in an oxide material, it changed its focus to zinc oxide. ZnO was known to be a semiconductor and therefore possibly related to electronics technology, and ZnO crystals were relatively easy to grow. Thus the Lander group was able to preserve itself and evolved within the Bell Labs structure by contriving a somewhat plausible alternative rationale for its existence.

My answer to 'now what' for 10 years

This was the framework within which I met chemist David Thomas and formed a working theory-experiment alliance that ultimately ranged over many different compound semiconductors. It was to provide me with unanticipated problems and paradoxes for years, to provide AT&T with a major knowledge base in compound semiconductors and semiconductor optics, and garnered David and me a joint "Oliver E. Buckley Solid State Physics Prize" from the American Physical Society in 1969. But at the time no one would have guessed that the combination of light and compound semiconductors would have the major technological future it now enjoys.

Leaving my sandbox

In 1968 I had run out of problems in condensed matter physics *to which my particular talents* seemed useful. I went to the Cavendish Laboratory of Cambridge University on a Guggenheim Fellowship for half a year hoping to find new interesting avenues, but found little for me. Returning from Cambridge University to Princeton University and my consultancy in the semiconductor group at Bell Labs, I ran into Robert G. Shulman, a chemist who was doing high-resolution Nuclear Magnetic Resonance experiments on Hemoglobin. He told me about the cooperative oxygen binding of the four iron atoms in the centers of the widely separated heme groups. An incredible panoply of physics techniques were being used to study the molecule. NMR, EPR, optical spectroscopy, resonance Raman scattering, X-ray structure studies, neutron scattering, Mossbauer Spectroscopy—all the clever experimental techniques of solid-state physics seemed relevant to Hb. For a

while hemoglobin was the physicist's 'hydrogen atom' for understanding how proteins function. Shulman wanted theoretical company to help interpret his NMR results, through which he hoped to understand the physical basis of the physiologically important cooperative oxygen binding. He knew the impact I had made through my interactions with Bell Labs (experimental) chemist David Thomas. So he made a strong effort to interest me in Hemoglobin problems, and the potential for such studies to move biology toward becoming a 'hard' science.

Knowing my talents, biology was a good place for me

In the mid 1970's I narrated an educational film about research in the interface between chemistry, physics, and biology. It contains a short clip in which chemist Linus Pauling responds to a question about how he chose a problem to work on. During this clip Pauling remarks 'I ask myself whether the problem I am considering is one to which *I* am likely to be able to make a contribution'. Even Pauling, who might well have received a Nobel Prize in Physiology and Medicine for the first understanding of an inheritable 'molecular disease' (except that he was already a double Nobelist in Chemistry and in Peace) found that a match between a problem and his talents significant. Knowing that a problem is important is not adequate reason to pursue it.

Hemoglobin provided me a simple entry from condensed matter physics to biological matter physics. How structure and low-lying excitations caused experimental physical properties was the name of the game, as was the case in much of solid-state physics. The one singular conceptual addition biology brought to the science was the notion of 'function'; that there is a small subset of properties that is of great importance to biology, and that evolutionary choices have shaped biological systems so that they function well. The term 'function' is peculiarly biological, occurring in biology and in applied sciences/engineering pursued to benefit humans, but not relevant to pure physics, pure chemistry, astronomy, or geology.

I interacted with the Shulman group for a couple of years on understanding the description of the interaction energy that caused the cooperativity in the equilibrium oxygen binding of Hemoglobin. Bell Labs was reasonably sympathetic to this venture. My consulting was moved from a semiconductor group to the biophysics group, with only the wry remark that Bell could scarcely raise my consulting fee that year since I was moving from an area where I was a world expert to an area where I knew nothing. It was a fair comment. We had some success in interpreting diverse experiments in a common framework. One of the then unheralded members of this group was Seiji Ogawa, who 20 years later became famous by using his expertise in NMR and Hb to invent Functional Magnetic Resonance Imaging (fMRI). fMRI makes images of how the brain processes information, i.e. how a brain *functions*.

The Shulman Bell Labs group next turned its efforts from the protein Hb to the nucleic acids called tRNA, where they could determine aspects of molecular secondary structure from NMR. Lacking any way to relate such experiments to functional questions, I lost interest in the group's experiments on tRNA. I did however attend *many* seminars from outside speakers who were describing the functional biological aspects of tRNA without knowing much about its structure. The one that still sticks in my memory 45 years later was by Herbert Weissbach on protein synthesis. Filled with far too many details for any physicist to remember, it included a funky movie of students playing the parts of amino acids, RNA, proteins, etc., culminating with a linked chain of amino students being generated while phosphate and tRNA students reeled off into oblivion. The overall impression I got was only that there seemed to be an extravagant waste of high-energy molecules in the processes of protein synthesis. The speaker, who was intent on describing a linear biochemical pathway for assembling a protein, did not mention my physicist's view of waste. Attending this two-hour talk was part of my penance for joining the biophysics group at Bell Labs. But it also presented me with a clear view of how biochemists viewed a complex problem.

At the same time I was giving my first biophysics course, designed for physics graduate students, at Princeton University. I spent an unreasonable amount of time on hemoglobin. Unfortunately hemoglobin is a poor introduction to problems of biology because its most obvious physics problem is a question of equilibrium. The essence of biology is the dynamics of a driven system often far from equilibrium. About five weeks into the term I sat down one evening determined to develop a theoretical treatment of *any* problem of biological dynamics. The sole precondition was that it needed to be handled at a level and fashion that required only knowledge of elementary quantum mechanics and rudimentary solid-state physics. I quickly realized that

from a physics viewpoint the simplest chemical reactions in biology were electron transfers with little nuclear motion and no rearrangements of chemical bonds. Early stages of photosynthesis and some of the important processes in oxidative phosphorylation were of this nature. So that evening I identified my topic for the next week and roughed out the solution to the electron-transfer rate problem.

Needing to coast a bit, I followed the electron transfer lectures by three weeks of standard biophysical lectures on cell membranes and the Hodgkin-Huxley equations for the propagation of nerve impulses along nerve cell axons. There was no originality or creativity at all in my presentation, but the lecture preparation began my grounding in neurobiology that would later prove invaluable.

For the last week of the course I made an effort to describe a dynamical (i.e. kinetic) problem involving tRNA. That class of molecules was chosen primarily because it was the only biological molecule system (other than hemoglobin) about which I knew *anything*. tRNA plays a central role in assembling amino acids into proteins, following the instructions on mRNA. Only a few minutes of physics thought was required to conclude that how this is accurately done is in part a problem in chemical dynamics.

Undeterred by my ignorance of biochemistry, I turned to the kinetic problems of *accurately* making proteins for the last week of lectures. Most of the biochemistry research and textbooks describing protein synthesis was based on a lock-and-key description of chemical reactions, with an incorrect reaction impossible because 'the wrong amino acid doesn't fit.' Understanding biochemistry was generally viewed as a problem of mapping out 'what happens'. What does *not* happen was not much thought about.

From a physics perspective, most biochemical reactions are possible at room temperature. There will only be different energies associated with similar but different reactions, and thus different Boltzmann factors determining kinetic rates. Discriminations are actually based on energy differences. 'The biochemist's 'A happens and B does not happen' should be replaced by 'A happens at a rate ~exp (-EA/kT) and B at rate ~ exp (-EB/kT).

The ratio of bad rates to good rates must be exp (-(EB –EA)/kT) where EB – EA the discrimination energy, a positive number. I managed to put together a few lectures showing that for accurate biosynthesis, a network of chemical reactions should not be pushed to operate too quickly. But the lectures involved no original constructs, and were classroom material, not research.

In the course of lecture preparation, I made a crude solid-state physics type estimate of the discrimination energy between two very similar amino acids, valine and isoleucine, that differ only in a single methyl group. I calculated ~1/50 for the maximum ability of any 'reasonable' isoleucine binding site to discriminate against erroneously using the smaller molecule valine. Unfortunately, the experimental number in biological protein synthesis (known at the time from the elegant work of physicist R. B. Loftfield (1963)) is about 1/3000. So much for my ability to carry estimation techniques from one field to another! As a warning against physicists' hubris, I described my obviously inadequate estimate to the class.

A possible problem–proofreading

After term ended, the issue still nagged me. After a month I realized that there *might be a real paradox.* My estimate could be roughly right. But Loftfield could also be correct, and the level of accuracy might *not* be determined by a simple discrimination energy. At the macroscopic level, a typist can proofread a document, and thus produce a final copy in which most of the errors in the original typing have been corrected. One way to explain the accuracy paradox was that cellular biology contains a way to proofread biochemical reactions at the molecular level, thus obtaining an accuracy (1/50)*(1/50) = ~1/2500 from an intrinsic fundamental accuracy of ~1/50. There might be a research problem in seeking to understand whether* in vivo* biochemistry carries out such proofreading, and we (i.e. biochemists) had simply not noticed the fact because we did not know to look.

An aside about the Overhauser effect

I owe much to many people, but particularly to my thesis professor Albert W. Overhauser. My identifying a research problem by finding a paradox to resolve comes straight from Al. But knowing Al had also made me think in a fundamental way about the meaning of his greatest research paper 'Polarization of Nuclei in

Metals' (A. W. Overhauser, 1953)). The subject matter is about as far from biology as physics can get. It is an astounding paper. I quote from his University of Chicago honorary Doctor of Science citation in 1979

> *"Overhauser proposed ideas of startling originality, so unusual that they initially took portions of the scientific community back, but of such depth and significance that they opened vast new areas of science."*

> *The consequences of this discovery—known as the Overhauser Effect—for nuclear magnetic resonance, and through nuclear magnetic resonance for chemistry, biology and high-energy physics have been enormous. The idea, which has also had very practical consequences, was so unexpected that it was originally resisted vehemently by the authorities in the field. Not until its existence was demonstrated experimentally by Slichter and Carver in 1953 was it fully accepted."*

Generalizations based on Overhauser's paper have had widespread impact on how NMR is used in determining molecular structure. The paper was the basis of his receiving the National Medal of Science. Kurt Wüthrich's 2002 Nobel Prize in chemistry was heavily based on Overhauser's understanding.

Allow me to make a brief digression into how Overhauser's paper could appear so wrong to almost everyone who heard about it. His theoretical prediction went against fundamental intuitions held by all physical scientists. Suppose someone tries to sell you a mug for making iced coffee in your microwave oven. You are told that this mundane mug has the property that most of the mug will (naturally) be heated by the microwaves, but as if by magic the coffee in the mug will become colder. This is exactly the snake-oil salesman pitch that Overhauser seemed to be making. He states that when a block of metal is exposed to strong microwaves, parts of the sample will become cooler.

Now, a refrigerator is a device that is driven by a power source, and heats up one part of the refrigerator while cooling another. So there is nothing impossible about the iced-coffee-in-microwave-oven mug if the mug were macroscopic and cleverly engineered. But the idea that every minute piece of lithium metal intrinsically behaves as a subtle refrigerator seemed absurd, even if (unlikely) you happened upon this line of general argument that reveals a fanciful possibility. But it is true.

Overhauser never wrote another research paper in this general area. (That this insight is not Overhauser's most cited paper underscores the futility of using the number of times a paper is cited in the literature as a measure of its importance to science.) But it so intrigued me at the time that I developed a simple understanding for myself, free from all the details of the physics of metals that dominated Overhauser's paper. Years later this understanding hinted that a biochemical relative of the Overhauser effect might explain how biology attains high accuracy, *if* the biochemical system also involved coupling to a chemical energy source. *The generalization was so metaphorical and removed from biochemistry that I did not mention it in the 1974 paper*, although it was extremely significant to my own thinking about proofreading.

Another way to increase fidelity is merely to wait—appropriately. When a chemical complex transiently forms, its probability of continuing to be bound decays exponentially, with a lifetime that depends on its binding energy. Transient complexes of a binding site with its preferred amino acid will have longer lifetimes than those with the incorrect amino acid. If as a 'Maxwell's Demon' one watched an unidentified amino acid bind to a binding site, but then waits a long time before using that amino acid, the 'wrong' amino acid would be exponentially less likely than the correct one to still be present and mistakenly used. The use of any such method requires waiting. Waiting requires knowing that time is passing, understanding the difference between the future and the past. To know this requires irreversibility and thus energy dissipation.

Almost there

Neither of these two insights described in detail how any particular biochemical system could 'proofread'. The only obvious common feature of these two viewpoints is that both concepts of how to get better accuracy from a fixed discrimination energy involve coupling the system to an energy source. Remember the Weissbach seminar! *Protein synthesis involves unexplained large amounts of biochemical energy consumption.* More pressing matters occupied my time for the month following this insight. But I now knew enough to be absolutely certain that when I could return to the problem, the details of how to proofread in protein synthesis would be obvious.

Proofreading lies hidden amongst the details of biochemistry

Within two months I had found plausible schemes for proofreading in protein synthesis based on known biochemical details, and described the kind of experiments that could be critical tests of whether proofreading occurs. For adding an amino acid to a growing protein, a particular complex containing one molecule of GTP, one molecule of tRNA, and one amino acid is used. The paradigm at the time would have described the correct biological stoichiometry between this GTP use and amino acids added as 1:1. Any measured departure from this integer ratio should be attributed to artifacts in the experimental study. My prediction was that the stoichiometry was not integer: that there should be slippage and the stoichiometry somewhat greater than 1:1 even for the addition of a correct amino acid, and that for adding an incorrect amino acid, the ratio should be large—certainly greater than 10:1. I had no idea of how to actually design a real experiment to test this idea.

The same basic reaction scheme for proofreading seemed to be present in DNA synthesis, in charging tRNA (i.e., linking an amino acid to its specific tRNA), and in assembling proteins. My research paper (J. Hopfield, 1974) described the reaction schemes of these three very different chemical processes as different ways of incorporating a simple unifying principle. While some kind of 'editing' descriptions had been understood for DNA synthesis prior to this 1974 paper, the general unifying energetic and kinetic aspects of proofreading and discrimination had not been recognized.

Even two years after publication the paper still did not have many molecular biology readers. That there were any readers at all was aided by the help that Bruce Alberts, (later to become president of the National Academy of Sciences) gave to the final manuscript. Noting that biochemists saw processes through details, and would reject an idea if any detail in its description were wrong, he carefully straightened out all my ignorant mistakes in chemical nomenclature. For this was the first paper I had ever written containing words like 'nucleoside' or 'synthetase' or 'isoleucine' or even 'GTP'.

I was occasionally asked to give biophysics seminars. One such a talk at Harvard Medical School in 1976 ended with a brief account of proofreading and my prediction about non-stoichiometry in protein synthesis. The second question at the end of the talk was from Robert C Thompson, a complete stranger to me. He merely asked 'Would you like to hear the results of such an experiment?' and proceeded to describe his experiments (not yet published) and the measured stoichiometry ratios that overwhelmingly supported my proofreading idea. He went on to describe how streptomycin killed bacteria by eliminating proofreading, leading to proteins with so many errors in them that the bacteria could not survive. (R.C. Thompson and P.Stone, 1977) **It was one of the biggest –and most delightful–surprises of my scientific career.**

The 1974 paper was important in my approach to biological problems, for it led me to think about the function of the structure of reaction *networks* in biology, rather than the function of the structure of the molecules themselves. A network could 'solve a problem' or have a function that was beyond the capability of a single molecule and a linear pathway. Six years later I was generalizing this view in thinking about networks of neurons rather than the properties of a single neuron.

Now what?? Finding A PROBLEM

I spent the long winter of 1977 at the Bohr Institute/Nordita in Copenhagen, as part of its sporadic but historical outreach toward biology. My nominal task was to run a broad 'modern biology for physicists' seminar series, with distinguished speakers from across Europe. I managed to recruit a powerful set of speakers with the help of the Bohr name. In that era, everyone still remembered Niels Bohr's interest in biology. Arriving in Copenhagen after a vacation in 1932, young theoretical physicist Max Delbruck rushed to the opening of the International Congress on Light Therapy where Niels Bohr was to give a lecture entitled 'Light and Life'. Bohr asked whether a deep explanation of life itself, and the philosophical ambiguities of how to interpret quantum mechanics, were intrinsically intertwined. The lecture was a transformative experience for Delbruck, setting him on the road from theoretical physics to a Nobel Prize in Physiology and Medicine. However, in 1977 my distinguished experts in biology emphasized what they understood, and skirted around all lacunae in their conceptual frameworks and research paradigms. They certainly did not describe biological science as being in need of help from (or hospitable to) theoretical physicists. No transformative experience

from hearing the 1977 lectures, for me or anyone else. Slightly disappointed, I returned to Princeton with no glimmer of a new problem for myself.

One can always construct variations and extensions on a theme of a previous paper, whether your own or that of others. The science research literature is overwhelmingly dominated by such work. But I was now looking for A PROBLEM, not a problem. The distinction between the two? In the early 1970's I served as the PhD supervisor of a highly imaginative but somewhat erratic chemistry student. After finishing a theoretical chemistry thesis with me, the student next became an experimental post-doc in the laboratory of James D. Watson at Cold Spring Harbor. I saw Jim about nine months later. Learning that my (former) student had not yet settled down to work on something, I made appropriate apologetic noises. Jim cut me off, remarking 'That's all right. When I first met Francis he was 35, and had not yet found a problem'. Now at that point in his career Crick had written 10 publishable papers, chiefly on x-ray crystallography, but what Watson meant was that Crick was working on problems, not A PROBLEM. Whereas what Watson already had already grasped, and which he was pitching to Crick, was the outlines of A PROBLEM that was the essence of all biology, the idea that knowing the structure of DNA could explain the chemical basis for biological inheritability. In mature hindsight, I was looking for A PROBLEM in biology for which my unusual background was unique and appropriate preparation.

Neuroscience/NRP

My office telephone rang very shortly before my morning traverse of the Princeton campus, from Jadwin Laboratory (physics) to Frick (chemistry) where in the fall of 1977 I was giving a course in physical biochemistry. That a chemistry department could permit me, who had never taken a course in either organic chemistry or biochemistry, to teach such a course was a consequence of University mismanagement of the interface between chemistry and biology at the highest administrative levels. That I would rashly attempt to do so was part of my looking for a significant new research project. Expecting a call from a tennis buddy, I hastily answered the telephone, only to find Francis O. Schmitt on the line. He said only that he ran the Neuroscience Research Program at MIT, would be transiting Princeton next Wednesday, and would greatly appreciate a half hour (no more) of my time. I had never heard either of Schmitt or the NRP, needed to head off for lecture, thought 'what's 30 minutes' and agreed to meet with no further conversation.

The following week, Francis O. Schmitt descended on me. He described an entity called the Neuroscience Research Program, which chiefly held small meetings in Boston attended by 20 regular members of the program and 20 visitors. The visitors were broadly chosen, but generally selected with an emphasis on the special topic under consideration at that particular meeting. Schmitt invited me to talk at the next meeting. I suggested that a physics graduate student in my biochemistry course, who had written a mathematical paper on neural coding, might give a talk of greater interest to the audience, a suggestion that Schmitt quickly rejected. I told him I knew nothing of neuroscience (a word Schmitt had coined some years earlier). He said that it didn't matter, "just speak on what interests you," so I talked about kinetic proofreading and the general issue of accuracy in the cellular biosynthesis of large molecules.

The audience—neurologists, neuroendocrinologists, psychologists, immunologists, electrophysiologists, neuroanatomists, biochemists—understood little of what I said. It didn't matter. Frank wanted to add a physicist member to the group, hoping to bring someone with diverse science experience to interact with his subject and perhaps help it to become more integrated and a more predictive science. Frank was a believer with a zealot's faith that somehow, sometime, science would be able to bridge the gaps between molecules, brains, minds, and behavior. He had gotten my name from Princeton relativist John A. Wheeler, who (for reasons that I have never grasped) had always been one of my staunch supporters. (Wheeler had also chaired the search committee that brought me to Princeton as professor of physics in 1964 because of my research in solid state physics.) It was a put-up job. Following Frank's leadership, the group voted to make me a member.

I was captivated by the talks at the meeting. How mind emerges from brain is to me the deepest question posed by our humanity. Definitely A PROBLEM. It was being pursued in narrow slices—as problems– by this NRP club of scientists with diverse talents and great enthusiasm. But it appeared to me that this group of scientists would never possibly address THE PROBLEM because the solution can be expressed only in an

appropriate mathematical language and structure. None then involved with the NRP moved easily in this sphere. So I joined the group, hoping to define, construct, or discover something I could usefully do in this field. The fact that Frank knew how to run a meeting with enough style and grace to attract the best of an international community into his orbit was also influential.

My basic education in neurobiology came from attending the semiannual NRP meetings, sitting next to world experts in their fields, who would patiently explain to me their interpretation of what was going on. Although Schmitt did his best to try to get experts to lecture broadly on their topics, and to describe neuroscience in an integrative fashion, he was generally defeated in this enterprise. So my introduction consisted of a set of disjoint expert views of experimental neuroscience, with interpretive commentary by an expert in some other corner of neuroscience. These commentaries were often impatient with the details studied by those in other subfields, but science is ever thus, and it was in any event not detail that I was seeking. I had tasked myself with finding an integrative view, of trying to somehow rise above the disjoint details of lectures ranging over primate neuroanatomy, insect flight behavior, electrophysiology in aplasia, learning in rat hippocampus, Alzheimer's disease, potassium channels, human language processing, ... to find a project to work on with the tools of theoretical physics.

Brains and machines 'compute' by following evolving state trajectories

Simple digital computers obtained their answers by starting at an initial state of the machine, implicitly described by a program and some data. They then change state again and again, at each tick of the computer clock, according to simple rules built into the hardware chips of the machine. Finally the state change stops. An end-state has been reached, where the rules generate no further state change. The answer to the programmed problem is the information now contained in a few specific memory registers.

Cellular automata are very special digital computers that had a brief moment in the sun in the late 1970's. They involve an array of equivalent 'cells', like the squares of a checkerboard (ignoring color). Each 'cell' has an internal state that changes in time in a deterministic fashion, according to rules that involve only the internal state of that 'cell' and the internal state of its neighboring 'cells'. All cells are equivalent, and change their internal states simultaneously. I first heard of cellular automata in reading about Conway's "game of life" in Scientific American, and I surmised that a generalization or modification of that basic idea might be useful in understanding how a brain operates. I speculated that if the cell state transition rules were made less rigidly structured, more like the network of synapse connections that provide inputs between neurons, and if the synchrony in time were relaxed to reflect neural delays in signal propagation and processing, it might be possible to bridge the conceptual chasm between digital computers and neural systems.

In the fall of 1978 I began playing with corruptions of the 'game of life' to make it slightly more like neurobiology, hoping to see it 'compute' by following the state trajectory to an answer. Unfortunately I was unable to carry out the mathematics necessary to follow the trajectory of the changing state for any such model. I needed to program a digital computer to simulate such a system, and to do computer experiments to gain insight into a diversity of such models.

It is difficult these days to imagine the primitive state of computers and computer laboratories in universities 37 years ago. Machines were slow, machine time was expensive, input to computers was chiefly through punched cards, output written on massive printers, TV-type display terminals rare. Computer power as measured by the number of transistors in a microprocessor, has followed "Moore's Law" for 50 years, roughly doubling every two years. That makes 18 doublings between 1978 and now. Thus I was stuck with computing power $(1/2)18 = 1/250,000$ of present day systems. With a very few notable exceptions (such as the AI lab at MIT) computers were used to produce numerical results from reliable programs and costly data. If you computer-simulated a model, it was because you had a well-based belief that the model was in good correspondence with reality. You did not experiment about a diversity of possible models—it was too expensive, a waste of a valuable resource. There was no emphasis on easy-to-program languages, and the computationally efficient languages were ungainly to use.

Princeton general computing and the computers of Princeton's high-energy physics group (the only departmental computer in physics) were run in a number-crunching mode. The idea of guessing the structure of a model, quickly and easily exploring the consequences of these guesses on a digital machine, and hoping to

find interesting models of neural activity evolution, was foreign to the computing facilities and environments of both Princeton and Bell Laboratories, where I also was affiliated.

Given my computing environment, I made little progress. The basic idea I wanted to pursue was that any computer, whether digital machine or a brain, operates by following a dynamical trajectory from a starting point (program and data) to an end-point, and that the trajectory needed stability against perturbations to get to the answer reliably in spite on noise and system imperfections. I did give an NRP talk on this idea of neurobiology as a system that computed with dynamical attractors. But there were neither computer simulations nor mathematics to give credence to this view. One Young Turk visitor came up to me afterward to tell me it was a beautiful talk but unfortunately had nothing to do with neurobiology. The others ignored it. I wryly note that my 2015 Swartz Prize (in computational neuroscience) from the Society of Neuroscience is really for this basic idea. But of course, the very existence of the term computational neuroscience implies that there are many mathematically sophisticated scientists now in the field, and these were an extreme rarity in 1979.

A better environment

In 1978 Harold Brown, the recently appointed Caltech president, resigned to become Secretary of Defense, and Caltech was again in the market for a physicist-president. They turned to Marvin (Murph) Goldberger, an eminent theoretical physicist who had chaired Princeton physics. Caltech, with Delbruck on the faculty of the Biology Division, had been making an effort to have more of a link between biology and physics. Goldberger had seen me struggling to do biological physics within the Princeton physics department. So during his honeymoon phase as Caltech president, he talked his faculty into offering me an endowed professorship jointly between Chemistry and Biology. The Caltech Physics Department, dominated by the mindsets of Murray Gell-Mann and Richard Feynman, had no interest in making a faculty appointment in such a direction.

What was the position of the Princeton Physics Department? I never gave a physics colloquium or physics seminar on either of my two most interesting biology-related papers (on kinetic proofreading and on biological electron transfer) written while I was in the Physics Department. The general attitude was that I was perhaps doing something interesting, but it involved too many details for Princeton Physics. Though Physics chair Val Fitch did, *sub rosa*, manage to appoint me to an endowed professorship.

When in October 1979 I went to see Val to tell him about the offer from Caltech, he said that sadly it would be best for both of us for me to go—best for me scientifically, and a simplification of his problem of departmental focus. There was no counteroffer.

February 1980. The quantum chemistry computing facility at Caltech was a splendid environment for trying out models. It supported multi-user real-time computing, with CRT displays and direct keyboard input, and no compilation delays. My research was a perversion of the intended purpose of this facility, but no one was looking. It rapidly became apparent that the previous year's speculation on relationship of brain computation to conventional cellular automata was useless.

A PROBLEM at last

It is surprisingly difficult to give up on a wrong idea that has been nurtured for a year. So rather than being totally abandoned, the cellular automata became perverted into a random quasi-neural network. The regular structure of cellular automata was abandoned in favor of randomly chosen connections. The complex set of logical rulers for state transitions was replaced with a rule inspired by biology. After a year of simulations and mathematics, I finally gave up on random networks. Instead, why not try a network with a particular structure chosen to accomplish some simple but profound task that neurobiology does rapidly, and that seems natural to biology but not to computers? The simplest such task in concept, and one that fit naturally into the basic computing paradigm of computing through dynamical system attractors, is associative memory.

Associative memory is reciprocal—seeing someone reminds you of their name (or at least did when I was younger), and hearing their name reminds you of what they look like. That fact can be expressed in network structure by making connections that are reciprocal. The mathematics of such networks is closely related to the mathematics of 'spin' systems that are responsible for all the complex forms of magnetism in solids. I

knew something of these systems through my connections with theoretical physics at Bell Labs. Suddenly there was a connection between neurobiology and physics systems I understood (thanks to a lifetime of interaction with P.W. Anderson). A month later I was writing a paper.

Publishing a paper

I had previously agreed to attend a symposium entitled "From Physics to Biology" of the Institut de la Vie at Versailles in the summer of 1981. This was an unusual gathering organized by Maurice Marois, a medical doctor who had dreams of strengthened connections between diverse scientists. He was persuasive with sponsors and flattering to Nobel Prize winners, and ran a posh, pretentious meeting with the conference lectures themselves in the Hall of Mirrors in the Palace of Versailles, and the speakers staying at the Trianon Palace Hotel next to the Chateau. I had happily (if slightly corruptly) accepted the invitation for such an all-expenses paid trip to Paris. I discarded my previously chosen topic in favor of a talk based on the recent work, making this talk at Versailles the first public talk on this subject. I have never met anyone who remembered hearing the talk.

The first manuscript I wrote on this research was a wide-ranging account of my recent research and its intellectual setting for the planned book of conference proceedings. When the organizers abandoned the book project, I set about converting my draft into a research article. I had two target audiences, physicists and neurobiologists, so thought immediately of publishing in PNAS. Neurobiologists read PNAS, and might see the article. And although few physicists regularly read PNAS in that era, at least PNAS was typically available in physics libraries. Not ideal, but the best I could come up with. As an Academy member I could publish such a paper without any review (this is no longer true, a sad commentary on aspects of science publishing and the promotion of originality). Distilling my material for PNAS was a challenge, for there was an absolute 5-page limit to article length, two audiences to address, and much to say.

Concerning the writing of non-fiction, Ernest Hemingway remarked,

> "If a writer of prose knows enough about what he is writing about he may omit things that he knows and the reader, if the writer is writing truly enough, will have a feeling of those things as strongly as though the writer had stated them." (E. Hemingway, 1932)

The PNAS length limitation forced me to be highly selective in what was said—and what was omitted. Had Hemingway been a physicist, he would have recognized the style. In hindsight, the omission of the almost obvious probably increased the impact of the paper. The unstated became an invitation for others to add to the subject, and thus encouraged a community of contributors to work on such network models. Successful science is always a community enterprise.

This 1982 PNAS paper is the first publication in which I use the word 'neuron'. It was to provide an entryway to working in neuroscience for many physicists and computer scientists. Further work connected these networks to many significant applications far beyond associative memory. It is the most cited paper I have ever written (6800 citations). Even AT&T was pleased (I had a part-time affiliation with Bell Labs all during this epoch), for the research also generated a very frequently referenced patent [http://www.google.com/patents/US4660166] for their patent pool, as well as strengthened links between neural biophysics and condensed matter physics at Bell Labs.

And now what?

In the introduction, I described the selection of what to work on as the most important factor in a career in research. The rest of this essay describes following the pathways of experience and investigation that led to two "Problems" well enough defined to become major research areas. In each case, there was a slow personal accumulation of the effects of deliberate steps and chance events that shaped the way that I viewed the world of science. This accumulation would shape my choice at the next fork in the pathway of possibilities.

"It's hindsight that makes a thing look inevitable. I wanted to write a memoir to explain how purely life depends on chance." (D. Hare, 2015). The playwright David Hare then went on to describe how he came to write his first play. He was, at the time, a young director with a strong political bent who had never conceived of himself as an author. But when an aspiring author failed to produce a promised script, Hare

stepped into the breach and wrote *'Brophy made good'* in four days, in order to have the needed new play for his troop of actors to rehearse the next week and perform the week after.

Reading this account, I felt 'that's me!'. One of my "plays" was ultimately entitled "A New Mechanism for Reducing Errors in Biosynthetic Processes Requiring High Specificity" (Hopfield, 1974). A long series of unpredictable events had led me from a childhood exposure to 'the world as physics' from my physicist parents, to condensed matter physics, to Cornell and Bell Labs, from there to the chemical physics of proteins, and finally to be teaching a Princeton course for which I had too little background. My need of lecture topics for my students loomed as acutely as Hare's need of a play for his actors. Previous chance garnering from biology lectures, a few random nuggets of knowledge from condensed matter physics, and I had what was necessary to begin writing a first draft of this "physicist's play". At no point prior to that teaching year had I the slightest notion of such a research direction.

What I have done in science relies entirely on experimental and theoretical studies by experts. I have a great respect for them, especially for those who are willing to attempt communication with someone who is not an expert in their field. I would only add that experts are good at *answering* questions. If you are brash enough, *ask your own.* Don't worry too much about how you found them.

References

E. Hemingway, *Death in the Afternoon* (1932).

A. W. Overhauser, Polarization of Nuclei in Metals, Physical Review **92**, 411-415 (1953).

T. R. Carver & C. P. Slichter, Polarization of Nuclear Spins in Metals, Physical Review **92**, 212-213 (1953).

J. J. Hopfield, Theory of the contribution of excitons to the complex dielectric constant of crystals, Phys. Rev. **112**, 1555-1567 (1958).

R. B. Loftfield, The frequency of errors in protein biosynthesis, Biochem. J. **89**, 82-92 (1963)

J. J. Hopfield, Kinetic Proofreading: A New Mechanism for Reducing Errors in Biosynthetic Processes Requiring High Specificity, Proc. Nat. Acad. Sci. USA **71**, 4135-4139 (1974).

R. C. Thompson & P. Stone, Proofreading of codon-anticodon interaction on ribosomes, Proc. Nat. Acad. Sci. USA **74**, 198-202 (1977).

J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proc. Nat. Acad. Sci. USA **79**, 2554-2558 (1982).

D. Hare, *The Guardian*, Review Section, Aug. 22 (2015).