Annu. Rev. Immunol. 1994. 12:1–62

THE WISDOM OF HINDSIGHT[1]

Melvin Cohn

Developmental Biology Laboratory, The Salk Institute, PO Box 85800, San Diego, California 92138-9216

KEY WORDS: concepts, immunology, theory, history, learning

Abstract

This essay is a highly personalized account of some of the important conceptual contributions to immunology. I have asked myself, "What were the ideas that caught my attention and how and by whom were they presented?" I have learned that most of what immunologists have called concepts deal with too small a slice of the subject. They are essentially inductive extrapolations from one experiment to a possible next step. Historically, these extrapolations extended over too narrow a chasm to account for the information available at the time. The result was that an extrapolation from one misleading observation could dominate and distort, for a significant time, the course of the field. It is also why there has been an inverse relationship between the clarity of a theory and its ease of acceptance by immunologists. Looking to the past, I have used two areas to illustrate the role of conceptualization: the self-nonself discrimination and the origin of the humoral repertoire. To illustrate all of this I have chosen as a cast of characters the founding fathers of immunology as we know it today. I hope that by taking this look into the rear view mirror our efforts will be guided in more productive ways. The take-home lesson is that we need to widen our horizon constantly to make more general concepts that then render the manipulation of the immune system more useful.

[1] Abbreviations: S/NS, self versus nonself; CN, copy number; iT or iB, T or B cell (initial state); eT or eB, receptor; MHC, major histocompatibility complex; CD, complementarity determining; FW, framework

0732--0582/94/0410--0001$05.00

PRE-AMBLING

Karl Popper spent the winter of 1966 here at the Salk Institute. There is no way that I can express the depth of his influence on me; I wrote an essay about this (1). We lunched together every day, just the two of us, and Socratically analyzed many ideas, among them the biology of expectation; in particular, how to formulate meaningful concepts in inherently complex areas. He put me in a good position to raise questions about the history of conceptualization because I now appreciate how much I learn by being wrong; I can change my mind when confronted with a rational argument, without the need to have the change appear to be purely semantic or to hope it will pass unnoticed. What must it be like to be a priest, general, bureaucrat, lawyer, medicine person, or politician who is never permitted to be wrong? No wonder they learn so slowly. I am grateful to be in a profession where, at least in my view, the realization of being wrong is equivalent to an increase in knowledge.

It is my belief that a meaningful history of immunology has yet to be written. By meaningful I mean one from which there is something to be learned about productive ways of doing science; something to guide our thinking process; something to change our choice of and approach to problems; something to teach us values. As we are all aware, every idea in science has roots in the past; it is onto these roots that new concepts are grafted (i.e. the historical process).

It might be argued that I should give up and accept that the history of immunology is as meaningless as its raconteurs make it out to be; an eclectic collection of household names and Nobel prizes uncoupled from the chain of reasoning required for *understanding*.

Almost anyone can make a major observation; however, it takes someone with a concept to fish it out of the ocean of minor observations. This is why, it might be argued, credit for a seminal discovery goes not to the one who makes it, but to the one who convinces succeeding generations of scientists to regard it as seminal. The contemporary generation is always tenaciously resistant to new ideas and change; in fact, being resistant to all but the smallest of changes is now a matter of survival.

I want to look at my own experience with a subject that has been occupying my thoughts for many years, namely, the role of conceptualization as an independent discipline specifically in immunology. I have found most immunologists distrust generalizing principles. The attitude, *never-theorize-when-you-can-do-an-experiment*, is not one found in the allied subjects of genetics, evolution, and molecular biology. While science is constructed from a collection of facts, a collection of facts does not make a science; they must be sorted, ordered, and brought together in

a theoretical framework. If it has any validity, a theory crystallizes a vast amount of experimental work and enables one to manipulate the information constructively.

In its recent history, immunology has advanced largely by sheer volume, complete with waste. Trials and errors are coupled to the lingering and unheralded death of that which is no longer heuristic or correct. The reluctance to discuss what was the cause of death means that nothing is learned. Immunology has fed parasitically off allied subjects, usually the purely technical, and, because it has provided no feedback, has influenced thinking in biology surprisingly little. In this respect, immunology as a science has not changed much over the past century; it has been insufficiently influenced by the need to understand in a coherent evolutionary context and overly motivated by the search, without worrying about understanding, for the fountain of miracle cures. In fact, not only has this approach guided the sources of research funds, but it has been elevated to the level of a philosophy by the tirelessly repeated argument that, after all, most cures have been arrived at by chance and with no understanding. In the face of this admittedly undeniable evidence, why waste time understanding? As a consequence, the molecular and cellular immunologist of today is no more or less driven by, or likely to search for, universal truths or coherence of context than the medical immunologists of the turn of the century.

The leading contributors of ideas in immunology have been empiricists in theoretician's clothing. This is best illustrated by Mitchison (2) who wrote, "Cohn makes an eloquent case for large theories. Personally I prefer small ones, perhaps mostly from habit. They may not be the currency of Nobel prizes, but they are the familiar coinage of everyday science, and they provide the excitement that keeps us going on dull afternoons. . . . An unattractive feature of large theories is that they force us to discard a great deal of perfectly respectable benchwork. . . . In contrast small theories invite us to confess our ignorance, by focusing attention on the gaps which they leave unexplained (2)."

Immunology does not have theoreticians as does a subject like physics; most of what we call theories are, indeed, small inductive extrapolations from an experimental observation ("small theories"); our subject is dominated by empiricism derived from its origins in medicine rather than biology. The limitation of small theories is their failure to bite off a large enough chunk of the subject in order that no one aberrant "fact" dominates the conclusion. I will illustrate how one such small theory, namely "the Landsteiner legacy," was singularly derouting for more than 50 years, and is still with us.

If we choose to explain one fact, most often, our choices of theory are

many; if we choose to explain more than one fact, our choices of theory become fewer and fewer as the number of relevant facts we choose to explain increases. In general, many theories explain few facts, few theories explain many facts, and finally, one or two theories explain a sufficiently large number of facts. Of course, any fact in contradiction either rules out the theory or itself becomes questionable as a fact. The fact questions the theory; conversely, the theory questions the "fact," that is, if it is a meaningful "large" enough theory. Mitchison has argued for one fact–one theory in order to fill our dull afternoons, whereas I am insisting on many facts–one theory in order to fill our gaps in understanding. When immunologists say they have developed a minimal ("small") theory, they mean, all too often, that either there are relevant facts the theory does not explain, or the theory is not reducible to the lower next level of mechanism, or it is not compatible with the higher next level of integrated function, or the theory is too vague to be disprovable, or it makes no predictions. All of this can be avoided by considering a big enough chunk of data, and this is much more than an afternoon's work—it is a profession.

Many immunologists would disagree and likely be offended by what I have just written. Therefore, let me pose the question for immunology rather than for immunologists: "Is there a role for a separate discipline called conceptualization in immunology and, if so, what are the bases on which it is to be built?" This is not a profound philosophical inquiry but, rather, a bread and butter question. Not implied here is the other extreme, *never-do-an-experiment-when-you-can-theorize*. Cartesian common sense or taste (the term I prefer) is an important factor in operating between the boundary conditions imposed by experiment and theory. In the spirit of this essay I try to discuss the question, using the wisdom of hindsight, by turning to history as I have lived it during a half century.

"MY VIEW OF THE WORLD": A PRÉCIS OF THE IMMUNE SYSTEM TODAY

The hindsight that I use is based on this précis of the immune system (see Refs. 3–8 for detailed analysis).

An immune system is characterized by two linked properties: (i) A somatic learning process to make a self-nonself (S/NS) discrimination; and (ii) a mechanism for determining the class of the response that optimally rids the target.

The first property deals with the recognitive aspect of the immune system; the second property with its effector function. These two properties are inseparably linked because evolution could not have selected for recognition without consequence. The necessity for this linkage is the evolu-

tionary selection pressure shaping the S/NS discrimination, the pathway of differentiation, and the size and origin of the repertoire. At no point in our thinking must this linkage be forgotten.

I recall that the S/NS discrimination is the *only* evolutionary selection pressure driving four specificity parameters:

1. the selectivity of the combining site of the antigen-receptor;
2. one specificity of combining site that is functional (signalling) per monomer of antigen-receptor;
3. one specificity of antigen-receptor per antigen-responsive cell; and
4. the activation signals (Signals [1]+[2]) operate within a single antigen-responsive cell.

These four parameters control the probability that induction by a nonself antigen will entrain an anti-self response, either because the secreted antibody to nonself cross-reacts with self [parameters 1 and 2 above], or because two antibodies would be induced, one anti-nonself and the other anti-self [parameters 3 and 4 above]. The induction of anti-self activity above a given threshold level results in autoimmunity. It is upon this anti-self threshold that evolution selects, thereby resulting in an adequate S/NS discrimination.

No selection pressure can operate to perfection. Selection against anti-self operates to a point where the frequency of autoimmunity no longer limits the survival of the species. At some point, the probability of being eaten by a predator becomes a stronger selection pressure than being debilitated by autoimmunity. Consequently all of these parameters have definable limits in that the selection pressure is on the integrated consequence of these four properties, which function to keep the average effect or level of anti-self activity acceptably low.

As regards the pathway of differentiation that permits a S/NS discrimination to be learned, antigen-responsive cells must be born with two pathways open to them, activation or inactivation, but with no effector function. We refer to this state as the initial or i-state (iB or iT). There would be no way to make a self-nonself (S/NS) discrimination if these initial or antigen-responsive cells were born as effectors. Further, it is only when in the i-state that a S/NS discrimination can be made. This consequence of the linkage between recognition and effector function will need to be faced over and over again as it is violated by many of the most popular interpretations of the immune system.

In order to respond to and rid a pathogen in a sufficiently short time, many antigen-responsive (i-state) cells must respond to it by becoming effectors. If the repertoire of *functionally different* specificities were $>10^{10}$, this would not be possible. A mouse has 10^8 B cells; if the *functional*

repertoire were 10^{10} and any one specificity were important, only 1 in 100 mice would contain it, and even that mouse could not respond before all but the most benign pathogens killed the host. In essence, above a certain functional size, the larger the repertoire, the longer the response time and the fewer the number of individuals that are protected. Therefore, the *functional* repertoire must be small (i.e. $\sim 10^5$) for both T- and B-cells and iterated in proportion to the size of the animal. The iterated minimum unit of primary protection is referred to as a *Protecton*. In sum, the concept of a Protecton arises from the requirement that a threshold level of effector activity be induced in a sufficiently short time. This is what links recognition to effector function.

What is the contribution of germline and somatic evolution to the functional repertoire of the Protecton? Again a generalizable consequence of the linkage between recognition and effector function must be highlighted. The functional repertoire is the summation of two repertoires, a high copy number (CN) repertoire of smaller size ($\sim 10^4$) and a low copy number (CN) repertoire of larger size ($\sim 10^5$). The high CN repertoire has $\sim 10^2$ copies per specificity; the low CN repertoire has ~ 1 copy per specificity. The high CN repertoire is derived by the unfolding of the evolutionarily selected germline as a first step. The low CN repertoire is derived by a second step of somatic diversification of the high CN repertoire. When an antigen is recognized by the high CN repertoire, the response is sufficiently prompt, but too few antigens are recognized by this small repertoire. When an antigen is recognized by the low CN repertoire, the response is too slow but few antigens will be missed. The two repertoires acting cooperatively provide an adequate recognitive potential as well as a sufficiently short response time. This is the fundamental structure characterizing the repertoires of all immune systems. In sum, the existence of a functional repertoire built upon this cooperative duality is driven by the limited number of choices that evolution had to create a "functional repertoire," given that recognition must result in an effective effector function.

Armed with these evolutionarily endowed parameters characterizing the Protecton, the immune system makes two key decisions upon encountering antigen (the target):

1. Is the antigen self or nonself?
2. If nonself, in which effector class(es) must the immune system respond to rid the target?

These necessary decision functions are the bases upon which one must view all theories of immune behavior. If a concept is incompatible with

either of these decision functions, it is not just meaningless; it is irrational, irrelevant, or absurd.

THE PATHWAY THAT I FOLLOWED

As this essay deals with my own experiences with the development of concepts, it would be helpful for the reader to get an inkling of the influences that operated on me. As I worked outside of the field of immunology for many years, I was able to watch the development of the field from another perspective. Further, I missed many of the influences that operated on most immunologists.

The First Era (1946–1949): A Brush with Immunology

In 1946 I became a predoctoral student of Dr. AM Pappenheimer, Jr. in the Department of Microbiology at New York University Medical School. Coming from a background in physics and physical chemistry and having spent five years as a soldier in the Pacific theatre and Japan, far away from academia, I found the idea of doing a doctorate in biochemistry terrifying. This department was unique in that it was in step with the elements that were to become the molecular biology of the future. As I was the only predoctoral student in the department, I became the apple of the eye for five awesome scientists, Drs. Colin Macleod, Ephriam Racker, Mark Adams, Alan Bernheimer, and Alvin Pappenheimer, Jr. or Pap, as he was affectionately known. Each expected me to know more than he did, an expectation that brings out the best in us.

Macleod was investigating the control of the pathogenesis of the pneumococcus. He would take a nonpathogenic "rough" strain of pneumococcus, mix it with the DNA derived from a pathogenic "smooth" strain of pneumococcus, and inject it into a mouse. Forty-eight hours later, just before the mouse died, he would culture "smooth" pneumococci from the blood. From his previous work with Avery, Macleod knew that the DNA transformed the rough strain in a type-specific manner. In his mind, the DNA specified directly or indirectly the structure of the capsular polysaccharide responsible for pathogenesis.

Racker, a master of the enzyme, made it clear that enzymes made carbohydrates, and Macleod's DNA could well be specifying the enzyme not the carbohydrate. Racker talked about Beadle's concept, one gene–one enzyme.

Bernheimer was deeply interested in the bacterial toxins, some of which today have been given a new role in pathogenesis as "superantigens" (a typical misnomer that characterizes immunology). They are the key factors in the pathogenesis of streptococci and staphylococci. Bernheimer wanted

to know what were their physical properties and mode of action. In this sense his interests directly meshed with those of Pappenheimer.

Pappenheimer had a passion for the "super-toxins"—diphtheria, tetanus, and botulinus. He concentrated on diphtheria toxin with the goal of understanding everything about it: physical properties, mode of action, conversion to toxoid, properties as antigen in several species including human. This latter was to be my introduction to immunology. I have written about my special debt to Pap, and I refer the reader to that article (9).

Adams was the chronicler of the phage group. He brought a unity to their activity, which, as it turned out, was the major initiating influence on the development of modern molecular biology. The most important point he brought to bear on our thinking was the 1943 Luria-Delbruck fluctuation analysis (10) showing that the mutational event preceded the selective pressure used to reveal it. Mutation was not induced by the revealing agent but rather was selected by it, implying the generalization that the ability to recognize precedes the stimulus (11). This he felt should give us pause in thinking about the template ("instructionist") theories (12–16), which, at the time, dominated immunological thinking; and well it did, in the minds of this group.

Immunology in the 1940s was, for me, a second order field. It was either a tool to study important biological molecules, or a tour de force problem in separating and characterizing a complicated mixture of molecules (immunoglobulins), or a field where the accepted conceptualizations ignored the remainder of biology. I felt incapable of formulating a conceptualization that would integrate immunology with what was already evident in other fields, particularly those occupying the NYU microbiologists, namely, genes and their products.

The Legacy Derived from Landsteiner: The Repertoire Is Transcendental ("Complete")

Landsteiner showed us, in Breughelian detail, that he could take almost any compound on his shelf, link it to an immunogenic carrier, and after suitable immunization, derive an antibody specific for it. In the chemically oriented world of immunology of the 1940s, this was a powerful tool, and one could now think about the stereochemistry of locks and keys in a very familiar way.

Nevertheless, a paradox was evident. The biologists were talking about genes and proteins, and genes were under evolutionary selection for the protein products they encoded. The immunologists were telling us that they could make a specific recognitive protein (an antibody) to any of a vast number of synthetic structures. Further, many different antibodies

could recognize the same determinant (i.e. degeneracy appeared to be high). How could the individual have accumulated the genes needed to encode a transcendental degenerate repertoire that recognized structures unlikely to be present in nature? This paradox is what made template theories so popular and acceptable; inexplicable observations tend to permit flights of fancy.

Unfortunately, this intuitive guess that evolution selected for a transcendental degenerate repertoire has remained dominant, even today. It has clouded thinking about the S/NS discrimination and the nature of the repertoire. How often I have read that clonal deletion is an absurdity because the repertoire is so degenerate that removal of anti-self will leave no antibody to react with nonself, or, that regulation via an idiotype network is unavoidable because a transcendental ("complete") repertoire cannot help but recognize itself. Or, of course, that the subunits of immunoglobulin are encoded by gene segments; they permit combinatorial rearrangement that yields a vast repertoire (what I have termed the "numbers racket"—79, p. 376). Few immunologists have dared question these self-evident truths.

What was missing is the relationship between recognitive potential and effector function. As I pointed out in *My View of the World*, evolution cannot select for recognition without consequence, and it is on this point that most thinking went awry. Given sufficient selection by an immunologist, it is not necessary to explain how a rare and exotic antibody specificity can eventually be isolated. It is necessary to explain what evolution selected upon, and in particular, the link between *repertoire* size and effective effector function. In order to function, a sufficient concentration of antibody must be produced in a short time; this is what limits the size of the functional repertoire, which must be small to be usable. I will return to the Landsteiner legacy over and over because it illustrates how often undisciplined intuition can be misleading. I hasten to stress that it was not Landsteiner himself who led us astray but the immunologists of his time who inductively extrapolated from what was on Landsteiner's shelf to infinity without worrying about what it takes to make an immune system functional. I refer to this view that the functional repertoire is transcendental ("complete") as the "Landsteiner legacy," with the above clarification in mind.

Soaring in the Clouds with Pauling

I would imagine that Pauling, a structural chemist, reasoned as follows: My good friend, Landsteiner, tells me that he can make an antibody to anything in the universe. As there are not enough genes in an individual, how might this be accomplished? Thusly, Pauling (12) posed the right

question, "what is the origin of the repertoire?" His best guess at the time was a template mechanism. Instead of questioning the equation: a finite number of genes (proteins) + an infinite number of antigens = an infinite number of antibodies; he accepted one absurdity and created another.

He made two assumptions based on the "rule of parsimony," first, that the Ig molecule of MW 150,000 was a single polypeptide chain, and second, that it had exactly two combining sites. These assumptions were not reasoned but arrived at by dead reckoning; the first was wrong, the second half wrong. As a consequence, his template theory required that the two combining sites be nonidentical, one anti-X, the other anti-Y. This shocked no one and for good reason: the S/NS discrimination was simply not a serious subject for thought; it could, therefore, be safely ignored. The molecular and the biological were not yet linked. If a significant proportion of Ig molecules had one site, anti-nonself, and the other, anti-self, there would be no way to make a self-nonself discrimination. Interestingly, this argument would not be accepted even today because most immunologists tacitly or unknowingly believe that the Ig molecule has a "sticky end" (multi-sited) (5).

There was no challenge to Pauling's theory of 1940 based on its total inability to cope with the S/NS discrimination until 1959, when Burnet (17, p. 61) wrote, "If antigenic determinants are needed to take a direct part in all antibody production, then there is no way of providing an interpretation of such phenomena as . . . immunological tolerance. . . ."

This period (1940–1959) was the era of template ("instructionist") theories. While most immunologists were instructionists, the theory had little influence on what they did; more interesting is the influence on what they didn't do. As a predoctoral student, I studied the papers of Haurowitz and Pauling (12–16), but I had no concept that permitted a choice between genes and templates.

A Startling Question Starts Me Thinking

In order to become a doctoral candidate, I had to pass examinations in French and German, physical chemistry, biochemistry, organic chemistry, physiology, microbiology, and history of science. One question still sticks in my mind (as does my answer).

As part of microbiology, I was asked to discuss the possible explanations for the finding that in the ABO blood group system of humans, the A individual expresses anti-B, the B individual expresses anti-A, the O individual expresses anti-A and anti-B, and the AB individual expresses neither. So thought provoking was this question that I found myself preoccupied with it, oblivious that it was an examination. I learned later that the question came, not surprisingly, from Pappenheimer.

There was no doubt that the antigens A and B were controlled by alleles, but what about the anti-A and anti-B? Were they under genetic regulation for expression such that the presence of A led to expression of anti-B, and of B, led to anti-A? And, if this were true, how was the phenotype of the O individual explained? After showing that no set of assumptions I could muster about linkage and control at the genetic level explained this finding, I concluded that the explanation was not at the genetic level but, instead, was a property of the responsiveness of the immune system. For example, an A individual might be unresponsive to antigens inducing an anti-A specificity but responsive to antigens inducing anti-B specificity. However, I was unable to formulate a more precise hypothesis. Nevertheless, I felt that I had ruled out all reasonable genetic interpretations of the relationship. Because I had read Landsteiner's book (18), I knew that this was a timely subject, but I could not reconstruct the other views, largely because there was nothing approaching a concept about them. In 1945, immunology was not the field to explore if you wanted to understand genetics or regulation.

My First Encounter with Burnet's Way of Thinking

Burnet & Fenner's article on "Genetics and Immunology" (19) appeared in 1948 when I was writing my thesis. It was so remarkable to me that they had focused on genes and antibodies in 1948 that I read the paper over and over trying to extract a coherent concept. To my surprise, they asked the same question that I had faced on my doctoral examination, namely, that of the relationship between the ABO blood groups and the anti-ABO agglutinins. They analyzed three then-published solutions to account for the characteristic expression of "natural human isoagglutinins."

1. The expression pattern of anti-ABO is apparent not real. The A individual absorbs anti-A making it undetectable, leaving anti-B in the blood; and similarly for the B individual.
2. The anti-A and anti-B agglutinins are induced by "an appropriate immunological stimulus."
3. Like the ABO antigens, the expression of the anti-ABO agglutinins are both "genetically determined."

I was sensitized to these three explanations because I had just invented the latter two for my examination and had rejected the third one. Let's consider why Burnet & Fenner opt for the third, albeit "an intellectually unsatisfying conclusion," that they state is at present, "the only available interpretation." They give us no idea as to why they make these qualifiers.

The first explanation was rejected because the evidence was against any anti-A binding to erythrocytes in the A individual.

The second explanation was rejected because of the "absence of evidence" for a source of A and B substances capable of immunizing the infant. Burnet & Fenner considered this sufficient grounds to make the hypothesis "untenable."

The third explanation was accepted by default, but they were "certain that it will eventually be replaced." Why and "replaced by what" was left to the reader.

There are two points of methodology to make:

1. There is a difference, to quote Forsdyke, between "absence of evidence" and "evidence of absence." The former is no argument for rejecting a theory; if anything, it should be an incentive to find the evidence. I appreciated this difference much later in my scientific career, and it has never ceased to be important as it comes up over and over again.
2. If a hypothesis is vague enough, it is neither testable nor heuristic. Both Burnet and Fenner were well-versed in genetics, and their analysis should have included a precise statement of what was "genetically determined."

A consideration of the S/NS discrimination is not mentioned by Burnet & Fenner as a factor in evaluating the three explanations, yet it occupies a large part of the remainder of their discussion. The first explanation might have been rejected a priori on this ground. They are ambiguous about the role of the S/NS discrimination, favoring a germline-encoded self marker theory to explain the ABO system, whereas a somatically learned S/NS discrimination is implied by the crucial Owen result that they cite. In discussing Owen's discovery of erythrocyte mosaics, they appear to have come close to an understanding, "cells 'foreign' to the host may be tolerated indefinitely provided they are implanted early in embryonic life" (19). However, in essence, this is merely a restatement of Owen's findings and stops short of an extrapolation to interpretation.

What was at the origin of this contradictory analysis? The answer is simple; at this point in their thinking Burnet & Fenner were essentially "instructionists," and this is what led them to propose a "self-marker" theory, to wit, any substance or cell that expresses the self-marker is nonimmunogenic. A small number of self-markers may vary from cell-type to cell-type in an individual. These self-markers are encoded in the germline on a "one self-marker–one gene" basis. In their view, the ABO system was an example of one such self-marker. "This concept of self-marker groupings as a necessity to allow immunological or non-immunological response as required, is the main novelty in the present review"

(19, p. 319), they comment. Therefore, the third explanation was favored as some genetic mechanism is required for a "self-marker" theory.

The importance of biting off a large enough chunk of data, and of being clear more than being right, is illustrated by the Burnet-Fenner contradiction. If their "self-marker" theory were correct, then the S/NS discrimination would be *germline encoded* (as it is for the defense mechanisms of invertebrates and plants), and their drawing of an "important implication" from Owen's observation would be an "absurdity" as this latter implies a *learned* S/NS discrimination. This is why in discussing the ABO system, a learned S/NS discrimination played no role; the ABO products are self-markers, and the absence of a response to them and their attached determinants is germline encoded.

As the vast majority of immunologists were "laissez-faire" instructionists, Burnet & Fenner's rejection of Pauling's direct template theory might have been important, but their argument was weak. They argued that if antigen were a template, then antibody synthesis would cease when antigen was ridded. They then cited evidence that antibody synthesis continued long after any antigen could possibly be present in the animal, but this remained controversial and unconvincing for a simple reason. All theories, on a priori grounds, require that induction of antibody cease when antigen is eliminated. Consequently, their argument rested on whether one could measure residual antibody secretion by end cells (plasmacytes) after induction had ceased, and this had to be beyond the experimental methodology of the time, thereby leaving in its wake a useless polemic.

What is surprising for me, in hindsight, is that the instructionist theory was not attacked on conceptual grounds, particularly since Burnet & Fenner placed their emphasis on the importance of explaining the S/NS discrimination. Simply put, Burnet & Fenner might have argued that there is no way to make a learned S/NS discrimination in an instructionist world, and therefore, that the theory is irrational or absurd. Why didn't they come to that conclusion? After all Burnet uses that as his main argument 10 years later (17, p. 61). The reason is that their competing theory was no less instructionist, as Talmage (20) pointed out, describing it as an "indirect template theory," as opposed to the "direct template theory." Burnet & Fenner (19) superimposed the self-marker hypothesis on the indirect template theory to account for a germline encoded (not learned) S/NS discrimination, but this concept could just as well have been superimposed on Pauling's direct template theory. If they had done that, the conceptual argument that template theories, direct or indirect, are incompatible with a S/NS discrimination would surely have emerged as they would have realized that a germline encoded S/NS discrimination in the form of a self-marker theory was untenable.

What Was Wrong?

Burnet was spared the criticism of contemporaries. This let him fumble uncorrected and reveals to us the arduous path of self-correction. As I discuss later, Burnet never did produce a theory of the S/NS discrimination, and this too seems odd, given his preoccupation with the subject. Eleven years later in 1959 he did produce a correct interpretation of the ABO system (17, p. 96). The extended version of the "indirect template theory (19)" was published in 1949 (21) when I was a postdoctoral student in Paris. Actually, I didn't read it until the late 1960s. Burnet remained with the indirect template theory for close to 10 years, until 1957; each successive elaboration of it (21, 22) only weakened the original formulation further. I would classify their "self-marker theory" as irrational even though I have treated it in the above discussion as erroneous; their "indirect template theory" was not a significant contribution to instructionist theories in general.

The Second Era (1950–1954): The Joy of Reductionist Molecular Biology

In 1947 Monod, a close friend of Pap, visited our laboratory. I was deeply influenced by my discussions with Monod (and passionately pushed by Pap) to carry out my postdoctoral work in Paris. I knew that before immunology could be opened up, the allied fields of biochemistry, genetics, and cell biology had to advance to a generalizable level; working with Monod on adaptive enzymes was one such avenue. I might, also, have been more influenced by the Burnet & Fenner paper (19) than I recall at this moment.

At that point in my career I was an experimentalist's experimentalist. I had carried out a set of tour de force characterizations of antigen-antibody interactions in several species including man that revealed universality and made it possible to analyze given proteins in complex mixtures of them. It was the use of this methodology that Monod and Pappenheimer felt would make important contributions to induced enzyme synthesis. I loved working at the bench and I did just that, spending long peaceful hours in deep concentration, following my protocol. The Pasteur Institute group was a joy to work with and once again I was with the best. I have written of this elsewhere (23); here only the immunological ties are relevant. This was a period where immunological methodology was key to our work. Using it, we were able to determine with uncanny accuracy the specific activity of β-galactosidase, its subunit size, and, with a crude molecular weight, that it was a tetramer. We further showed that it was synthesized de novo from amino acids following induction (24). For all of this immunology was a reagent, not a serious subject for study.

Watson visited our lab in 1952 bringing the solution to the structure of DNA before it was published; with little guessing, we were convinced that the principles of base pairing extended to *transcription of RNA* and that *RNA was translated into protein* using a code that would someday be elucidated. (Italic indicates modern terminology.) This was a time when the Brachet formulation was on our blackboard, DNA → RNA → protein. In our Paris laboratory we held these truths to be self-evident and universal.

My only link to immunology during the years 1950–1954 was the next door laboratory of Pierre Grabar in which Jacques Oudin and Alan Bussard worked. I occasionally went to their seminars and marveled at the new techniques of gel diffusion and electrophoresis. During these years, immunology was, for me, a science of successful recipes, and it was the recipe that I admired the most. Being a devoted experimentalist, and a damned good one, I only accepted ideas that were close to the data, and it would never have occurred to me to consider seriously, much less publish, a theory that extended itself much beyond the experiment. I played ideas like a winning poker hand, close to the chest.

One late rainy afternoon in the autumn of 1953, Jacques Monod, Martin Pollock, and I were relaxing in the laboratory sipping cognac and discussing the dismal failure of a set of "on-the-side" experiments that I had carried out to see whether any anti-β-galactoside antibodies could be found that possessed enzymatic activity. This led Monod to comment that it was time to introduce genes into the analysis of the immune system when, out of nowhere, Pollock made a remark that startled me. He said, "Why don't immunologists consider the possibility that antigen selects for cells?" His comment had much the same electrifying effect on me that Watson's informal lab seminar to our group on DNA structure had had a year earlier. Monod, in his characteristic way, treated Pollock's idea as self-evident, but for me, obvious or not, it was exciting. I decided that if ever I were to return to immunology, this idea was worth pursuing experimentally. It has always impressed me that years later, when I reminded Pollock of his suggestion, he had no recollection of having made it.

The Third Era (1955–1967): The Return to Immunology

I returned to the States in late 1954 (Department of Microbiology, Washington University, St. Louis) and continued my work on induced enzyme synthesis. However, Pollock's suggestion kept returning to my thoughts, but what experiment would be critical? If antigen selected cells thereby amplifying the antibody response, then the repertoire of a single cell had to be small compared to the total. How small? I knew that I was dealing with a diploid cell and if the answer were "one cell–one antibody" there would have to be a mechanism to silence one chromosome or segment of

it. This did not seem impossible as such a phenomenon was known for the X chromosome. If the answer were two antibodies per cell, no special mechanism of exclusion was needed but a great deal of useless antibody would flood the system. The relationship of the number of antibodies produced per cell to the S/NS discrimination was to be understood by us much later (4, 25, 32). Also, I liked the idea of one cell–one antibody because the inactivation of cells anti-self would permit a modest beginning to the understanding of the S/NS discrimination. The more I thought about it, the more enamored I became of the idea.

At the regular midwestern meeting of the "phage group," in Urbana in 1955, organized by Luria and Speigelman, I met Ed Lennox who was interested in doing something in immunology. We discussed the idea that one cell made one antibody and came up with two approaches to measuring the antibody secreted by single cells, both derived from the bacteriophage methodology. First, we would try to visualize the antibody secreted by a single cell as a plaque; second, we would measure phage neutralization by a single secreting plasmacyte in a microdrop. We divided the work: Lennox investigated the microdrop approach and I, the plaque assay. I immunized rabbits in the footpad with sheep erythrocytes, and at various times made a cell suspension from the popliteal lymph node that I plated with sheep erythrocytes and complement in a soft agar layer analogous to the phage plaque assay. No plaques were visible; yet, in suspension culture, the cells secreted large amounts of hemolytic antibody. The failure was quickly pinpointed; the agar was a very anti-complementary, and we had no clue as to how to solve that problem. The two solutions that appeared years later were to find another gelling medium, and to neutralize the anti-complementary activity of agar with DEAE-dextran; neither the reagents nor the know-how was available to us in 1955.

Fortunately, Lennox's preliminary studies were encouraging, and we turned to phage neutralization in microdrops and developed a unique methodology for the assay of many cells. For all of us working on this problem, the field was new. I was turning to cellular problems in immunology after many years of work in bacterial genetics and induced enzyme synthesis; for Lennox, it was a change from theoretical physics to bacteriophage genetics to cellular immunology; for Attardi, the pathway was from medicine to cellular immunology; and for Horibata, it was from bacterial physiology to cellular immunology. The stories of each of the players in the puzzle is of great interest in itself, but that is not my goal here. As my laboratory was in St. Louis and Lennox's laboratory was in Urbana, we set up a natural double blind experiment. The experiment was carried out in St. Louis where the single cells in microdrops were visually scored and then incubated with bacteriophage, which were then plated for

counting. The numbered Petri dishes, hundreds of them, were shipped by Greyhound bus to Urbana, where they were counted. Then the decoded protocol was used to sort out the results independently in St. Louis and Urbana, as an internal error check.

Before describing the results, I would like to make two comments. First, we were floored to discover that someone else in the world was trying to answer the same question. We should not have been surprised that this was Burnet's group and that Nossal and Lederberg were associated with it. Second, we were confirmed and even prejudiced unispecific clonalists. We had little doubt as to the outcome. Consequently, we confidently pushed the experiment by repeatedly and heavily immunizing the rabbits, selecting those with popliteal lymph nodes as big as golf balls. We pushed the immunization because we knew that the probability of expression of an "Ig-gene" had to be low if "unispecific clonality" were to be achieved. The demonstration of this required that we saturate the expression potential, and our guess was that this required maximum immunization.

That Fateful Experiment

Our result (26–31) was that roughly 5% of the total secreting plasmacytes, anti-reference antigens, were double producers. The finding was shocking, depressingly so, because we confidently expected to find none. Given that our reference antigens were the major antigenic load, the fact that 95% were single producers established clonality, but why wasn't the result "picture perfect," that is, no doubles? We were driven to extremes to eliminate artifact, and in the end, we learned that there is no such thing as "zero" in biology. This experiment, of course, said nothing about the initial population of antigen-responsive cells (iB) except that some small proportion were double producers and under fierce antigenic selection, these could be revealed experimentally. We were to learn eventually that this population is the target of the evolutionary selection pressure due to the S/NS discrimination and cannot be reduced to zero (4). All that should have been in question at the time is "At what level are the doubles?" However, before discussing this key point, it would be valuable to add this personal note.

I believe today both on experimental and conceptual grounds that our findings were absolutely correct. Nevertheless, our failure to find perfect "unispecific clonality" was the most costly result of our scientific careers. The bandwagon was rolling, instructionism was being swept away, and "unispecific clonality," with no real conceptual context, was filling the vacuum. We were forced by our peers, who were running as a pack, to become outsiders (eventually, a real plus) in the face of the failure of other investigators to find "doubles." In 1967, at the Cold Spring Harbor

meeting, our study was treated as totally discredited, as due to artifact and therefore of no positive value. In reality, this is still true today (e.g. 81, pp. 159–161). Needless to say, the suggested and implied artifacts were, in fact, eliminated by our controls; the origin of the low level of doubles either was due to an artifact both unknown to everyone and specific to our study, or else it would someday be understood as real. The latter proved true.

Our experiments (26–31) were simply more thorough and complete than any others, picking up the background class that proved the rule; unispecific clonality obtained, but it is the consequence of evolutionary selection for a S/NS discrimination that reduces the doubles. Both the immunological community and even we, ourselves, treated these studies as anathema. As we had no conceptual framework onto which unispecific clonality might have been mapped, we had no way to evaluate and defend our results. Consequently, we accepted the verdict of the community that an unknown artifact was one possible explanation.

As I read these papers (26–31) today, they are classics in this field, experimentally too far ahead of their time, and therefore, misleading in the period 1959–1967 when no understanding of unispecific clonality in a proper biological context existed. The intellectual residue of instructionism was dominant, with the result that the role of the S/NS discrimination was not yet a major factor in thinking. I was well aware by 1972 that *clonal selection is not an independent theory* but is derivative as a corollary of the denial both of instructionism and of a germline encoded S/NS discrimination (32, pp. 8–9) Yet it took until 1987 (25, p. 684) to state clearly that the driving selection pressure for "one cell-one antibody" was the S/NS discrimination, and to quantitate its effect by calculating the acceptable level of doubles upon which evolution selects. In hindsight, that understanding this point took us so long is embarrassing.

In the section "My View of the World," I pointed out that the S/NS discrimination is the only selection pressure driving four properties, one of which is haplotype exclusion; the consequence of this is one cell–one antibody. I also pointed out that this selection pressure operated to reduce autoimmunity to an acceptable level by keeping the anti-self response during an anti-nonself response below an effective level. Clearly it cannot operate to reduce the anti-self response to zero and, therefore, cannot reduce the doubles to zero. We also understand today the two mechanisms of haplotype exclusion, the one used by mouse and human (4), the other by birds and rabbits (7). In neither case could these mechanisms eliminate total doubles much below 5% of total antigen-responsive cells. Molecular and cellular biology confirm our earlier studies that a small proportion of antigen-responsive cells are double producers (4, 7, 25), and their existence

is key to understanding the selection pressure maintaining unispecific clonality. The general rule is that a selection pressure cannot operate if there is nothing upon which to select (i.e. the doubles).

What Was Our Real Error?

Our experiments were a classic example of seeking to find support for our prejudice, in this case, unispecific clonality. We were not experimentally deciding between two equally valid theories. Consequently, we had no framework that could deal with an unexpected answer. Given this, we probably never should have carried out this experiment. Before it could be made meaningful we needed to understand that unispecific clonality is driven by the S/NS discrimination. No evolutionary selection pressure can operate to perfection, which is unselectable. We might have seen this as the reason that autoimmunity exists. If we had understood then what we understand now, the debates would have had a quite different flavor, and immunology would have had a little push in the right direction. Also, as a confirmed experimentalist, I was simply not mentally prepared to play the dual role of theorist. But, for me, this experience eventually became a major driving force compelling me to put understanding before experimenting.

What Was Happening Conceptually in Immunology While We Were Buried in This Fateful Experiment? The Role of the Big Three

Most historians trace "selectionist" theory from Jerne (33), to Talmage (20) and Burnet (34), and correctly so as this is how it was introduced into the immunological community. However, this was not quite the route I followed. In recounting this period I have been especially careful to reread the original papers because I find that many who have reviewed this history were, in fact, rewriting it.

Jerne's paper (33) appeared in 1955 when I was at Washington University in St. Louis. Jerne proposed that "antigen is solely a selective carrier of spontaneously circulating antibody to a system of cells which can reproduce this antibody (33)." "Spontaneously circulating antibody," or "natural antibody" as it was referred to at the time, was considered by Jerne (33) to be derived from the continuously varied, random synthesis and secretion of globulin molecules, which generated an enormous variety of different combining sites, anti-self and anti-nonself. The anti-self is filtered out in fetal life; the anti-nonself survives in the adult. Those globulin molecules that react with nonself were then template- or self-replicated to increase their concentration.

I immediately presented this paper at our department journal club. The

response was like mine; it seemed inconceivable that one could propose a self-replicating protein, given what we knew about DNA structure and genetics in 1955. Neither was there any need to convince me that template theories were ruled out. What was needed was a plausible selectionist theory, and it had to begin with a receptor on cells not in solution. Lennox and I were already at that point in our thinking. I was rather surprised that Delbruck submitted Jerne's paper to *Publications of the National Academy of Science*, but immunology was only an eccentricity and could be allowed to run counter to the rest of biology.

Jerne's proposal (33) had an impact in that it oriented the thinking of both Talmage (20) and Burnet (34) by providing them with the seeds of a competing view to that of Pauling (12). Unfortunately, I did not read the two important papers that Talmage (20) and Burnet (34) published in 1957 until 1959 when they were referenced in Burnet's book on clonal selection (17). I had missed Talmage's paper (20) because the discussion central to antigenic selection of cells was not indicated by the title "Allergy and Immunology" and Burnet's paper (34), because the *Australian Journal of Science* was not among the journals in our university library. These were the days before *Current Contents* or "medline" searches, and those immunologists who did read these papers were not sufficiently fired to spread the gospel at the meetings I attended.

Talmage (20) made three points concerning Jerne's theory:

1. "As Jerne has indicated, the natural selection theory gives a simpler and more definitive explanation for the absence of auto-antibodies. If these proteins or their synthesizing units are eliminated during fetal life, they will not be available later for selection and multiplication."
2. "While Ehrlich's and Jerne's theories agree in considering antibody formation as a process of natural selection, they differ considerably in the mechanism by which this selection occurs . . . thus, according to Jerne, the basis of replication is an extra cellular protein, whereas according to Ehrlich a replica is made of some intrinsic cellular unit. The latter hypothesis is preferable. . . ."
3. "The process of natural selection requires the selective multiplication of a few species out of a diverse population. As a working hypothesis it is tempting to consider that one of the multiplying units in the antibody response is the cell itself . . . only those cells are selected for multiplication whose synthesized product has affinity for the antigen injected. *This would have the disadvantage of requiring a different species of cell for each species of protein produced* . . ." (my emphasis).

Burnet (34) introduced the term "clonal selection," which has since been

used ambiguously. To some it means only selection of cells by antigen. To others it implies, in addition, one cell–one antibody. I use the term "unispecific clonality" to imply one cell–one antibody. Agreeing with Talmage (20) Burnet (34) added an important stress, namely the need to explain the S/NS discrimination. He, too, cites "the great virtue of the Jerne hypothesis is that it provides an approach to this alternative method of recognizing self from not self (34)." He correctly rejected Jerne's theory (33) on the grounds that a self-replicating protein was unlikely. Both Talmage (20) and Burnet (34) seem to distinguish between a multi-producing clone and a uniproducing clone. In Burnet's (34) words, "Each type of pattern [the recognitive combining site of an antibody] is a specific product of a clone of mesenchymal cells, and it is the essence of the hypothesis that each cell automatically has available on its surface representative reactive sites equivalent to those of the globulin they produce."

While Talmage (20) gives us no reason for unispecific, as opposed to oligospecific clonality, Burnet (34) argues that "the theory requires at some stage in early embryonic development . . . a 'randomization' of the coding responsible for part of the specification of gamma globulin molecules, so that after several cell generations . . . there are specifications in the genomes for virtually every variant that can exist. . . ." Thus, Burnet tries to answer the right question but inverts the logic in his answer. The statement "All men are mortal" does not require that all mortals be men.

If the generation of diversity were due to "randomization" of genes, then unispecific clonality might result (depending on a great many additional assumptions Burnet would need to have added). However, if unispecific clonality obtains, it does not mean that the generation of diversity is due to somatic "randomization"; that is only one of several pathways to unispecific clonality as the subsequent discussions on "germline" versus "somatic" theories illustrated. In the end, Burnet's argument (34) was wrong; unispecific clonality is driven by the necessity to make a S/NS discrimination, not by any requirement of the generator of diversity (3, 4, 25, 32). This error of logic was perpetuated unchallenged by immunologists for the succeeding three decades up to the present.

Lastly, is it really true that Jerne's theory (33) gave us a simpler and more definitive explanation of the S/NS discrimination as Talmage (20) and Burnet (34) emphasize? "Natural" antibodies according to the theory (33) are effector molecules generated and secreted antigen-independently as anti-self and anti-nonself. Two a priori laws of the S/NS discrimination are that (i) there is no way to make the S/NS discrimination at the level of effector function, and (ii) effector antibody cannot be used to regulate responsiveness because it is tied obligatorily to destructive ridding mech-

anisms (3). Burnet (34) was vaguely aware of a problem and added one of the missing assumptions, namely, that the repertoire of "natural" antibody had to be generated by a "big bang" during fetal life and then stabilized so that it remained unchanged for the life of the animal (sounds like the T cell 1993?). As suggested by Jerne (33) anti-self would be filtered out [without activating any destructive effector mechanisms!] at which point the "natural" antibody generator would have to cease production leaving its already secreted product, a complete repertoire of anti-nonself, to function for the life of the animal. To say the least, this made Burnet's addition (34) questionable, but it was still far from the heart of the problem. As a matter of principle, the S/NS discrimination cannot be determined at the level of effector function.

Regulation by effector antibody (in the absence of concomitant destructive effector function) is a conceptual error that was repeated during the idiotype network era because, at the time, no one took seriously the question of why this class of theory (33) for the S/NS discrimination had to be wrong, in principle (3). Finally, a proposal for a mechanism of tolerance is only one part of the S/NS discrimination, as I will point out later.

What was important during this period (1955–1960) was the realization that even for the immune system the recognitive element must precede the stimulus (11), the basic tenet of selective theories. Antigen (the stimulus) does not instruct the formation of the corresponding antibody (the recognitive element); it selects for it. Immunology could now be tied by a thread to the mainstream of biology. One might have supposed that the 1943 experiment of Luria & Delbruck (10), the 1949 experiment of Newcombe (35), and the 1952 experiment of Lederberg & Lederberg (36) were finally being understood by immunologists. In fact, immunologists arrived at their conclusion independently and without definable conceptual roots. No one appeared to be aware of these fundamental investigations based on sound conceptualization (10, 35, 36), as they were never cited.

What Was the Key Contribution of the Triumvirate?

Without negating the Landsteiner legacy of a transcendental repertoire, Jerne (33), Talmage (20) and Burnet (34) took the emphasis away from repertoires and put it on cells and their regulation. This was psychologically important as it permitted an integration of the molecular and the biological. Either one alone easily leads to absurdity. Yet, at this stage, no one had actually integrated the thinking about molecules and the cells that make them. In the end, most of what the triumvirate achieved was the establishment of a language with which the proper units of immunology might better be manipulated.

Was the Next Step Expected?

Two groundbreaking papers by Lederberg (37) and Talmage (38) appeared in the same issue of *Science*, 19 June 1959 (which will be a collector's item for immunologists one day). Both had a profound influence on my thinking. I am always impressed that these two papers are never reproduced in the anthologies of milestones in immunology collected by historians to guide students.

Lederberg (37) mapped immunology onto the emerging molecular biology of the time. His analysis is a model for balanced thinking and taste in science. Lederberg introduces the terms "instruction" and "selection" to describe the two theories for the role of antigen. He, too, is unclear as to the function of unispecific clonality, tending to explain it alogically, as did Burnet, as a consequence of the generator of diversity. As a result, it is remarkable to me that Lederberg was the first person to give us a theory of the S/NS discrimination.

A theory of the S/NS discrimination must answer three questions:

1. What interactions of antigen with the antigen-responsive cell lead to its inactivation (tolerance)?
2. What interactions of antigen with the antigen-responsive cell lead to its activation (induction)?
3. How is the decision between induction and tolerance arrived at and how is it maintained throughout life?

Lederberg (37) argued that cells are born tolerizable-only. An interaction between antigen and its receptor leads to inactivation (Signal [1]). After an adequate period without encountering antigen, the cells differentiate to an inducible-only state in which interaction with antigen leads to induction to many effectors (Signal [2]). The translation of his theory into our present language (3) would be that cells are born in the i-state (initial, antigen-responsive state) and interpret the interaction with antigen as inactivation (Signal [1]). They then differentiate in an antigen-independent step to the e-state (effectors), not distinguishable from Lederberg's inducible-only antigen-responsive cell that interprets the interaction with antigen as divide (Signal [2]). The decision between tolerance and induction is made by the state of differentiation of the B cell and the "persistence" versus "transcience" property of self versus nonself antigens. In Lederberg's language, "If an antigen is introduced prior to the maturation of any antibody-forming cell, the hypersensitivity of such cells, while still immature, to an antigen-antibody reaction will eliminate specific cell types as they arise by mutation, thereby inducing apparent tolerance to that antigen. After the dissipation of the antigen, reactivity should return as

soon as one new mutant cell has arisen and matured (37)." Lederberg thus pinpointed correctly the only difference between self and nonself antigens, namely, that self is present when antigen-responsive cells are born and it persists, whereas nonself is transient, appearing after these cells have matured to effectors, and is eliminated by them. His theory turned out to be wrong, and for a priori reasons; there would be no way to control the mutants of anti-nonself cells to anti-self cells, a lethal situation (32). Nevertheless, it was a giant step forward, and subsequently, it became incorporated into a larger context, but in a way that Lederberg could not have arrived at in 1959. It was not until 1983 when Langman and I (3, 4, 6, 39, 40) answered the question, "How does the immune system get started?", that the larger context became evident.

Talmage (38) focused our attention on two important points, both of which directly questioned for the first time the Landsteiner legacy. That he did this is remarkable to me.

The first challenge derived from the fact that the humoral immune system can distinguish a vast number of antigens without requiring an equally vast number of antibodies. Talmage illustrated this by considering how a repertoire of T antibodies can "distinguish" a family of M antigens. He argued that the total number of antigens, M, "distinguishable" by T antibodies is the combinatorial, ${}_{T}C_{q}$, where q is the average number of different antibodies reactive with a given antigen. Using his example, if each antigen is seen by 3 antibodies, a repertoire of 5 antibodies could "distinguish" 10 antigens (i.e. ${}_{5}C_{3} = 10$). Thus he argued validly that the repertoire could be small. But is it? The answer cannot come solely from a consideration of specificity. The problem is twofold: First, an effective effector response (destructive and ridding) in a sufficiently short time must result from the "recognition" of antigen; and second, the combining site of antibody must have sufficient selectivity ("specificity") to "distinguish" self from nonself so as to avoid autoimmunity concomitant with the ridding response to nonself. The repertoire of T functionally different antibodies must "effectively recognize" as well as "distinguish" antigens.

There are three types of relationships between antibodies and antigens (viewed as being composed of linked antigenic determinants):

1. A given antibody molecule can react with a family of structurally related but distinguishable antigenic determinants (i.e. cross-reactivity).
2. A given antigenic determinant can react with a family of distinguishable antibody molecules (i.e. degeneracy).
3. An antigen is a collection of distinct antigenic determinants recognized independently by antibodies that can be described as functionally different.

Protecton theory (4) tells us that the third case is the one of major consequence, but Talmage (38) deals essentially with degeneracy (case 2). The first case, that of cross-reactivity, is of limited functional significance for the evolved immune system; any antibody that cross-reacts with a nonself and a self determinant is treated as anti-self. Degeneracy is of minor functional significance because any family of antibodies distinguishable by sequence that reacts with the same determinant is a single functional antibody (i.e. they would be indistinguishable at the functional level as effector molecules). The third case is central to an aggregative effector function because, for a monomer, $\geqslant 3$ different determinants must be bound to make a matrix aggregate. Antibodies recognize antigens, determinant by determinant. The Talmage calculation applies to this case also. As a rough approximation, because we do not know the distribution function relating determinants to antigens, T antibodies will effectively recognize ${}_{T}C_{q}$ antigens. While a small number of antibodies can both effectively *recognize* and *distinguish* a large number of antigens, this statement does not permit a calculation relating specificity to repertoire size because the two are not directly related. Specificity is driven by the S/NS discrimination, whereas the size of the functional repertoire is determined by the factors responsible for generating an effective and safe response in a short enough time given a limited number of iB-cells per ml of animal (see Ref. 4 for detailed discussion).

In his second challenge Talmage (38) introduces the seemingly obvious but nonetheless crucial point that, "the reactions between antigens and antisera are strongly dependent on concentration and have as well sharp thresholds below which no reaction can be detected." Using a mean value of the binding constant of antibody and a reasonable distribution of values around the mean, a threshold concentration of antibody for effective function can be determined. Talmage thus pinpointed a fundamental axiom upon which Protecton theory (4, 25) was much later to be built.

Talmage (38) is not clear as to why he posed and tried to clarify these two prescient points. Lederberg (37) gives us his answer, "it would embarrass a theory of cellular selection only if it [the size of the antibody repertoire] is large compared with the number of potential antibody-forming cells in the organism." Here it is 1993 and immunologists, still burdened by the Landsteiner legacy, are multiplying gene segments to calculate repertoire sizes $> 10^{10}$ to describe a murine immune system with 10^{8} B cells per animal. Lederberg (and presumably Talmage) realized this absurdity (i.e. "the numbers racket") in 1959. It is clear that the emphasis on cells instead of molecules was beginning to pay off by making one ask such questions as "what does an understanding of the S/NS discrimination entail?" and "what is the relationship between recognitive potential and effector

function?" For me, Talmage and Lederberg began to point the way out; Jerne and Burnet, while making important contributions, remained mired in the Landsteiner legacy.

All of this said, we are left to wonder why Talmage and/or Lederberg did not derive Protecton theory (4, 25). In fact, it might more appropriately be asked, what took us so long, some 30 years, to see where Lederberg (37) and Talmage (38) were leading; the moment should have been ripe to tie recognition to effector function. Why were we such unprepared minds in the 1970s? Today, I am certain that we were simply too inhibited to chew off a big enough chunk, and thus we stayed too close to the comfortable, albeit dull, afternoon theories of the type favored by Mitchison (2).

What Was Solved and What Was Unsolved?

At this point it is important to make a critical evaluation of the status of thinking. The thrust was the development of "selective" theories of immune responsiveness. The major conceptual advance was the defining of the unit of selection as a cell expressing a receptor identical to the one that it would secrete upon being "selected" (induced). However, one fundamental property of an immune system was ignored during this period. No one considered the obligatory linkage between recognitive potential and effector function. For example, Jerne (33) considered an aggregative interaction between "natural" antibody and antigen that had as a consequence (effector function) the cellular uptake and replication of the "natural" antibody, but he never considered that a concentration of "natural" antibody high enough to carry out that effector function would be high enough to carry out other effector functions, which are destructive. He repeated the same error 20 years later when he proposed regulation via an idiotype network. As I have said, Talmage (38) gives us food for thought by correctly posing the question of the size of the repertoire, and Lederberg (37) points out correctly why this question is important; yet, neither understood that it was essential to link recognitive potential to effector function. As a consequence, Talmage is essentially preoccupied with the degeneracy of antibodies (a second order problem) and "distinguishing" one antigen from another (as a definition of specificity), but neither of these factors limits the size of the functioning repertoire. In contrast Burnet (34) and Lederberg (37) using an inverted logic (see earlier) arrive at a repertoire composed solely of unique or single copy mutants (a nonfunctional immune system). If a given pathogen were recognized by one mutant cell per mouse (10 mls) and stimulated to divide at 0.5 days per division, it would take $\geqslant 18$ days to produce an effective concentration (100 ng/ml) of antibody; if the pathogen were recognized by 10 mutant

cells it would take $\geqslant 8$ days. Both of these response times to a bacterial pathogen are too slow. A single pneumococcus will kill a mouse in <5 days. Finally, it is absurd to have to conclude that the immune system of a human is superior to that of a mouse in protecting against primary infection because a human is bigger. Using that argument, elephants should have close to perfect immune systems.

Talmage (41) chides the immunologists of the 1960s for being refractory to new ideas. This has always been true of the response to all new concepts, alas! However, in defense of those refractory immunologists, the new ideas were not presented with the keenest of arguments or a broad consideration of other biological fields. Jerne (33) not only ignored the molecular biology of the day, he also provided the most superficial of explanations of the S/NS discrimination. Talmage (20) correctly put the emphasis on selection of cells by antigen, a major conceptual advance for immunologists but a resolved concept for microbial geneticists (10, 35, 36). He stressed that the "multiplication of cells is required rather than multiplication of subcellular units (20)." The question of cellular versus subcellular selection by inducers of "adaptive enzymes" (referred to as the "plasmagene" or "duplicon" controversy) had also been settled by microbiologists as being selection of cells (i.e. the postulated "plasmagene/duplicon" turned out to be a cell) (42). Thus far unispecific clonality was not in question; only selectionism versus instructionism was involved. Burnet (34) introduces "clonal selection" as an extension of Talmage (20). It is defined as one cell producing one antibody being selected by antigen. The origin of unispecific clonality is incorrectly assumed to be "that . . . cells . . . can be regarded as belonging to clones which have arisen as a result of somatic mutation." Burnet (34) had to be aware that cells are diploid. The correct conclusion can and is sometimes arrived at by the wrong reasoning. The unispecific antigen-responsive cell (haplotype exclusion) is selected by the necessity to make a S/NS discrimination, not by any property of the generator of diversity. Burnet (34) tells us how anti-self cells might be eliminated, but not how anti-nonself cells might be induced or how the choice between a self and nonself antigen is made by a mature immune system. Thus, he never produced a theory of the S/NS discrimination. Lederberg (37), who is never referred to as part of this period even in hindsight (e.g. see Ref. 43), made clear finally that the fundamental question was "selection versus instruction" and gave us a precise example of what a theory of the S/NS discrimination entailed. Whether the cell selected by antigen had the potential to express more than one antibody was a separate and further question, once "selection" was established. However, Lederberg (37) (surprisingly, to me) accepted Burnet's incorrect argument that "unispecific clonality"

is required by the generator of diversity (e.g. somatic mutation). It took many years for Langman and myself to map "unispecific clonality" correctly onto the S/NS discrimination (4, 25, 32) and, thereby, to place evolutionarily selectable limits on the proportion of double producers.

Science and Hero Worship

We must be careful not to treat our heroes as do their parishioners (43), but rather we must constantly evaluate their steps forward in terms of their real contribution to our understanding. While the emotional hindsights characterizing the biographies written by nonhistorians are delightful reading, much more is learned by the criticisms of a thoughtful colleague than by the adoration of the Magi. This comment in no way need detract from the inspirational leadership of great people.

According to Sexton (44), Burnet felt he deserved two Nobel prizes: one for his insight on the origin of tolerance, for which he did receive the prize, the other for clonal selection theory, for which he did not receive the prize. Burnet had a point: the Nobel Prize committee sometimes gives the prize to the right person for the wrong reason.

To quote Sexton: "Burnet always regarded this hypothesis [clonal selection] as the finest of his theoretical accomplishments, and one that, to his mind at least, was deserving of a Nobel prize. In a letter of congratulations to Jerne on his award of the Nobel prize in medicine and physiology for 1984, he was to write: 'I have often thought that you and I should have had a joint award for putting antibody production on the right track rather than the one I shared with Medawar. Anyway we are both now on the list'.

Apart from Burnet's own belief that the clonal selection theory merited a Nobel prize more than, or at least the same as, his other major immunological hypothesis [origin of tolerance], there is no doubt that his micro evolutionary explanation of the adaptive nature of antibody production heralded in a new era in immunology" (44, p. 140).

To quote Burnet, "In that year (1949), Fenner and I published a second edition of an Institute monograph on *The Production of Antibodies.* In this there was the first clear recognition that the differentiation of self from not-self was very important in immunology and that, to a large extent, it was developed in birds and mammals during embryonic life. That book is long out of print and has become a minor collectors' piece because of a certain prediction made on p. 103: 'If, in embryonic life, expendable cells from a genetically distinct race are implanted and established, no antibody response should develop against the foreign cell antigen when the animal takes on independent existence' " (44, pp. 136).

While this shows that Burnet had a healthy and normal level of conceit

(today we call it self-esteem), more to the point would be the placing of his comments in historical perspective. Putting aside the limitations to using a Nobel prize as a measure of true creativity or even achievement, Owen (45) was aware, as were many other immunologists, that his finding implied that the S/NS discrimination is learned during fetal life (i.e. when the immune system is immature). In any case, it is none too insightful to make an inductive extrapolation to all antigens from Owen's findings in chimeras with one antigen, erythrocytes. I view Owen's experimental finding taken at face value, not Burnet's restatement of it, quoted above, as the major advance. Why?

Although Burnet deserves credit for highlighting the Owen discovery, neither the indirect template theory nor the self-marker theory developed in the monograph *The Production of Antibodies* (21), cited by Burnet, implied the prediction (44, p. 136) quoted above; an experiment did. On the contrary, the self-marker theory implied a *germline* encoded (not a *learned*) S/NS discrimination that is totally *independent* of the state of development of the animal's immune system. Where was the reasoning that led to the "prediction made on page 103" that required the conclusion that the only difference between self and nonself was learned as the ever presence of self and the transcience of nonself? Proper argumentation might have led Burnet to a theory of a learned S/NS discrimination, which, despite his emphasis on the question, he never produced.

Burnet did play a major role in defining and popularizing "clonal selection theory," but as I pointed out, he also introduced numerous confusions. As he was uncorrected by his colleagues, all of whom tried to live in his light, the theory did not grow in his hands or those of his associates, either with respect to correcting the errors or to integrating it with the molecular level findings.

Burnet influenced my thinking by insisting correctly that Darwinian evolutionary principles be used in analyzing the immune system, and by shifting attention from molecules to cells, thereby giving immunology the balance it desperately needed at the time.

The Fourth Era (1968–1980): Coming To Grips with Self

The S/NS discrimination needed conceptualization if we were to make the next step, particularly because the Lederberg model (37) of the S/NS discrimination had been experimentally disproven and we needed one to fill its place.

The Origin of a "Two Signal" Model of the S/NS Discrimination

Peter Bretscher joined my laboratory in 1967; he was, by way of background, an X-ray crystallographer who had a passionate curiosity about

immunology as well as a clear, crisp way of analyzing complex problems. This made the difficult question of the S/NS discrimination an ideal one to answer; we were aware that it had to be analyzed correctly if a next step was to be made in immunology.

We had two prior formulations, those of Lederberg (37) and Forsdyke (46). As I have said, Lederberg (37), by example, illustrated the three questions to be answered if we were to develop a theory of the S/NS discrimination. We knew that the theory of Lederberg no longer fit the data (47, 48); it was some years before I realized that the theory was wrong on a priori grounds, because mutation to anti-self in an inducible-only cell would be lethal and such mutations are unavoidable (32). Forsdyke (46) gave us the barest of hints as to what a competing theory would entail. Using an analogy with the coincidence circuitry of a liquid scintillation counter, a light flash due to radioactive decay in the scintillation vial would be read simultaneously by two photocells and be counted. However, a spontaneous discharge in a single photo cell due to any background "noise" effect would be ignored. To this analogy, Forsdyke (46) added two points: (i) "foreign determinants...are unlikely to be constantly present... self determinants are likely to be constantly present," and (ii) "activation of one site might mean 'self,' and activation of both sites simultaneously might mean 'non-self' (or vice versa)." The first point had been made previously by Lederberg (37); the second point was the seed of a new idea.

Forsdyke (46) suggested that self determinants combining with two distinguishable antibody sites of identical specificity in close proximity (a doublet) lead to deletion of those anti-self cells. Those cells (anti-nonself) that do not encounter antigen secrete spontaneously (antigen-independent) "natural" antibody. When nonself antigen encounters the immune system, the "natural" antibody blocks the determinants on the antigen so that it reacts with only one of the cellular sites and an immune response is initiated. The proposed mechanism had to be wrong, even though in one form or another it was reintroduced independently by several investigators over the years. Once again, the linkage between recognition and effector function was ignored. Forsdyke's (46) unique idea was that two separable and distinguishable signals to the cell are required to separate inactivation by self from activation by nonself. Lederberg (37), I recall, postulated two signals initiated by identical interactions separated in time by a step of differentiation so that they were interpreted differently. Forsdyke (46) proposed a situation where an interaction leading to a doublet results in inactivation, whereas one leading to a singlet leads to activation, in any given cell. While this, too, had to be wrong on a priori grounds, a two-signal mechanism distinguishing self from nonself by each antigen-responsive cell was required, and Bretscher and I set out to develop just such a theory. It

took until 1989 for Langman (3) to give us a formal argument as to why two signals are required for a S/NS discrimination.

How Did the "Two-Signal" Theory Develop?

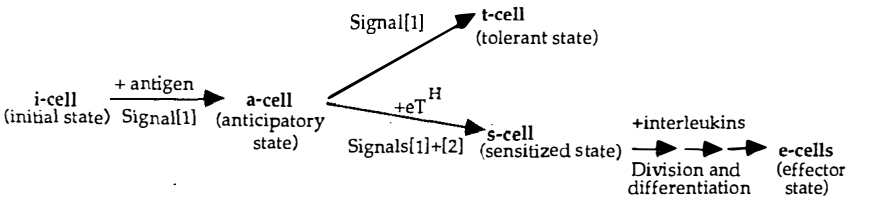
The "two signal model" (as it was later dubbed by Möller) had a rocky intellectual history; but, as formulated today, it is highly likely to be correct. In essence, there is at present no validly competing model. In order to develop its pathway, I state its elements as we understand them today and then show how many wrong steps and filling in of incompletenesses were necessary to arrive at the present model. This is how we learn; peer criticism would have hastened the process, but the concept was essentially ignored until recently when it gained a dubious notoriety.

There are several essential elements:

1. The S/NS discrimination is a learned process; it cannot be germline encoded for immune systems as it is for the defense mechanisms of invertebrates and plants. This is the first point to examine when presented with a theory of the S/NS discrimination.
2. Stem cells give rise to unispecific antigen-responsive cells. We refer to these cells as initial state or i-cells (iB and iT). They have no effector function at this stage; they receive signals, they do not send them. In order to be referred to as "unispecific," the i-cell population must be over 95% haplotype excluded. Lastly, i-cells must have two pathways open to them, one resulting in effectors (induction) and one resulting in death (tolerance); *it is only when in the i-state that a S/NS decision can be made.*
3. The interaction of the antigen-receptor (BAr for B cells, TAr for T cells) with antigen results in a signal to the i-cell (iB/iT) that is referred to as Signal [1]. If the i-cell receives no other signal, it is inactivated irreversibly (deleted, killed) with a half life of roughly 0.5 day. This is the tolerance pathway.
4. If an i-cell receiving Signal [1] (i.e. on the pathway to death) interacts with an effector T-helper cell (eT^H) that delivers Signal [2], then the i-cell is activated and put under the control of interleukins that drive it to division and differentiation to effectors (eB or eT). The regulation of the pathway is transferred from a cell-cell interaction (eT^H-Ag-T/iB) to a cell-interleukin (sB/sT-interleukin) interaction (see schema below). The function of the cell-cell interaction is to regulate the S/NS discrimination. If, in order to monitor mutants to anti-self, a cell-cell interaction were required at every division, the response would be too slow to protect against an average pathogen. Consequently, the control is transferred to the more rapid cell-interleukin interaction,

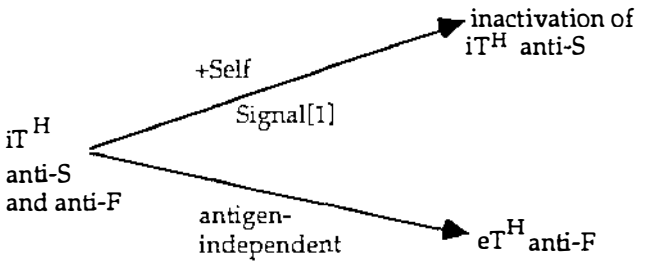
which cannot regulate the anti-self mutants that arise. They accumulate as long as there is an ongoing response to the foreign pathogen. When the latter is eliminated, the anti-self response ceases. The cells anti-self are eliminated either because they are dead-end effectors or because they are returned to the i-state where they undergo a S/NS discrimination and are tolerized. Because they are long-range messengers, interleukins are involved not in the S/NS decision, but rather in the expansion of an effective effector response. The S/NS decision requires short-range communication as exemplified by that mediated by cell surface–cell surface interactions such as restrictive recognition of antigen.

Schematically the inductive pathway is:



The decision between the two pathways inactivation (tolerance) and sensitization to responsiveness to interleukins depends upon the insufficiency or sufficiency of effector T-helpers (eT^H). During fetal life, when i-cells are being generated, there is an insufficiency of eT^H. Consequently, all i-cells anti-Self (anti-S) are deleted by interaction with Self(S). When the immune system matures and a sufficiency of eT^H is generated, which is anti-nonself (F = foreign), the presence of a foreign substance now induces an effector response.

5. How does the immune system get started; where does the first eT^H come from? As the eT^H-cell is obligatory to induction, this theory requires that there be an antigen-independent pathway generating eT^H. This pathway is schematized below:



We proposed that this antigen-independent pathway to eT^H has two characteristics:

1. The time that it takes to differentiate to effectors (eT^H) in the absence of antigenic encounter must be sufficiently long compared to the time that it takes to find a self antigen and be inactivated (tolerized). The proportion of iT^H anti-S that slips through and becomes eT^H anti-S will be a function of the ratio of these two times. As the level of eT^H anti-S must be kept acceptably low, the overall rate,

$$iT^H \xrightarrow[\text{independent}]{\text{antigen}} eT^H,$$

must be sufficiently slow to allow time for the adequate deletion of the iT^H anti-S, yet it must be sufficiently fast to permit a priming level of eT^H that enables an adequate anti-nonself response to be established.
2. The antigen-dependent pathway (i.e.

$$iT^H \xrightarrow[+\text{ antigen}]{eT^H} \text{many } eT^H)$$

is characteristic of a response to a foreign antigen (F) and must be fast by comparison with the antigen-independent pathway.

It is this proposal as to where the first eT^H-cells come from that carries the imprint of Lederberg (37), referred to earlier.

How did we arrive at this formulation? To discuss the elements in the progression of a concept, I must clarify several terms that I use.

"Unresponsiveness" will be used when referring to an experimental finding. "Tolerance" will be used when extrapolating the observation of "unresponsiveness" to a concept of how the S/NS discrimination is determined.

"Unresponsiveness" is of two forms: positive or dominant unresponsiveness is referred to as "suppression," and negative or recessive unresponsiveness is referred to as "paralysis." Both terms are used when referring to experiment.

In 1968, the initial formulation (47) of the theory was sufficiently correct to be able to extract from an enormous and confused experimental literature the three empirical relationships to be explained.

1. The establishment of unresponsiveness in a mature immune system is antigen-concentration dependent whereas the maintenance of unresponsiveness (or the establishment in an embryonic immune system) is antigen-concentration independent, of course, when above a given threshold (the Mitchison phenomenon).
2. For an antigen to be immunogenic, at least two determinants must be

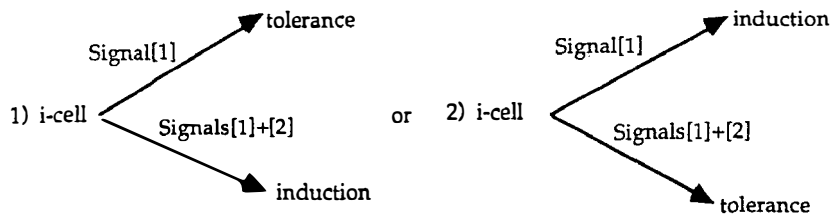
present on the antigen to which the animal is not tolerant (i.e. foreign (F) determinants) (the Benacerraf phenomenon). We often referred to this as the hapten-carrier effect.

3. There is competition between unresponsiveness and induction at the level of the antigen-responsive cell (i-cell) (the Weigle phenomenon).

Our theory dealt adequately with the Benacerraf and the Weigle phenomena but rather poorly with the Mitchison phenomenon. It took time to become suspicious as no theory seemed adequate. Today we know why. The observed unresponsiveness at low and high concentrations of a distinctly foreign antigen was due to suppression, not paralysis. The phenomenon of suppression cannot be extrapolated to the mechanism of the S/NS discrimination; as formally demonstrated in an elegant paper of Langman (49). Suppression, and hence the Mitchison phenomenon, is involved in the regulation of the class of the response, not the S/NS discrimination. This is a good example of a theory putting an observation into its correct context, and of the importance of biting off a large enough chunk initially to avoid being misled by one datum.

As an aside, high zone unresponsiveness due to paralysis (negative or recessive unresponsiveness), if convincingly demonstrated, would have far reaching consequences. High zone unresponsiveness was interpreted by us as due to the saturation of antigen-receptors on the eT^H and iT/iB cells blocking interaction between them via an antigen bridge—hence, tolerance (32, 48, 50). If, as believed today, the TAr can recognize only peptide-MHC, then this explanation is not possible, and paralysis established in the high zone would be unexplained; it would become a contradiction to this assumption concerning TAr. specificity. A cell-cell interaction dependent on recognition of peptide-MHC cannot be blocked by the precursor protein prior to processing because the antigen-specific receptors involved in the cell-cell interaction are postulated to be blind to it.

Forsdyke (46) had made Bretscher and myself aware of the need to decide which of the two pathways required Signal [2], and we rejected the Signal [2] requirement for the tolerance pathway. The two choices were:



The second choice was initially rejected on experimental grounds (i.e. recognition of the carrier is required for induction, not unresponsiveness). Later the second choice was rejected on the a priori grounds (49, 51) that suppression cannot regulate the S/NS discrimination.

If I may be permitted a reflection, all too often a dominant, albeit correct, idea narrows one's horizons, shutting out the ability to place an observation rejected in that framework into another framework where it is uniquely illuminating. I was so certain that choice 1, above, was correct that I failed to appreciate two important observations on establishing unresponsiveness described at the Brook Lodge Symposium in 1968 (52). These were the establishing of "ultra low dose unresponsiveness" and "carrier effects" in establishing unresponsiveness. As "ultra low dose unresponsiveness" implied the induction of a shut-off mechanism (i.e. suppression) and "carrier effects" implied that the role of Signal [2] was to inactivate, not activate (choice 2), I put them aside in search of an explanation, which I fully expected to be artifactual. Although it was correct that these observations suggesting choice 2 were not germaine to the S/NS discrimination, they did demonstrate suppression without anyone's being aware of it, and long before it was definitively pinpointed by Gershon & Kondo (53). While even today I am annoyed with myself for such cycloptic thinking, I did eventually learn to be wary of an overly closed mind. "Suppression" has been uniformly misinterpreted as a mechanism of tolerance; however, it will play a major role in understanding how the class of the effector response is regulated, and that is today's most important immunological problem. The misplaced pigeonholing of suppression as part of the mechanism of the S/NS discrimination even today is why I insist on the distinction between unresponsiveness and tolerance in our discussions (51).

The Mechanism of Signal [2]

What I want to develop here is the evolution of our understanding of the mechanism of Signal [2]. The recent acceptance of the "two-signal model" I earlier characterized as "dubious" because today those who accept the model treat the mechanism of Signal [2] too casually. The mechanism linking recognition of the carrier to the delivery of Signal [2] must confine it essentially to the i-cell receiving Signal [1]. The delivery of Signal [2] to innocent bystanders must be minimized. A short range Signal [2] is one part of the specificity package that I outlined in the section "My View of the World."

The requirement for an adequate S/NS discrimination is that two determinants on the antigen be recognized, one by the i-cell (Signal [1] and the other by what we referred to in 1968 as "carrier antibody." This was

postulated to be of a special class, IgT, the function of which was to initiate Signal [2].

Our 1968 proposal (47) for Signal [2] was conceptually wrong. The receptor on the i-cell was pictured to undergo one conformational change on interacting with antigen (Signal [1]) and another one upon interaction with two particles of antigen bound together by carrier antibody (Signal [2]) (i.e. the stretched conformation). This proposal that the BAr could undergo two distinct conformational changes corresponding to the two signals was conceptually wrong because it is central to the theory that Signal [1] be included in the inductive event (Signals [1]+[2]) and not be an alternate state. The reason is that the inclusion ensures no i-cell would be inducible that, in principle, could not have been tolerized; evolution would have been very receptive of that argument. Our early realization of this error paid off in understanding.

In 1970, we (48) corrected this error by postulating that a conformational change in the "carrier antibody" was read as Signal [2] by the i-cell that had bound antigen (i.e. receiving Signal [1]). This is linked or associative recognition of antigen. Thus, Signals [1] and [2] were separated but delivered simultaneously or in close sequence to the same i-cell.

In 1971, we (50, 54) made the "carrier antibody" cytophilic for an effector cell (a carrier specific cell). This latter delivered Signal [2] by associative recognition of antigen bound to the i-cell. It seemed unrealistic at the time to propose a cell-cell antigen-driven interaction between two unispecific cells because, for an effective response, they were expected to take too long a time to find each other. A secreted antibody of a special class seemed to be the most likely solution to making the rate of the interaction sufficiently rapid to account for an immune response. Further, we postulated that "carrier antibody" was made by a T cell that we referred to as a "cooperating T-cell." Mitchison named this cell the "helper T cell," a name that stuck, unfortunately. The reason I opposed the use of the term "helper" was twofold. First, the helper concept was that the T cell "concentrated," "focused," or "trapped" antigen so that B cells of low affinity could be induced. This is absurd as there is no value to inducing B cells that secrete antibody of too low affinity to function effectively in solution. The T cell directly or indirectly had to mediate a signaling function (54, p. 547, footnote 18). It has always astonished me that this obvious argument was never accepted. Second, Signal [2] is obligatory for induction under this model; the "helper" concept was that the signalling role, if it existed, would be facultative (frequently normally bypassed). Three beautiful papers analyzing this stage in our thinking were written by Bretscher (50, 55, 56).

By 1973, it was clear to us that we were dealing with an effector T-

helper(eT^H)—iT/iB interaction that did not involve secreted antibody for which I had used the symbol IgT in 1971 (54). We viewed the cell-cell interaction as mediated by recognition of two determinants on the antigen, one by the i-cell (Signal [1]) and the other by the effector T-helper (eT^H) that delivered Signal [2] via a synaptic structure formed between the two cells. The synapse was activated to signal by the surface receptor (IgT)-antigen interaction. This structure was the precursor model to what we refer to today as restrictive recognition of antigen. Today, the signalling synapse is between the restricting element (MHC) on the target (acceptor) and CD4 (donor) linked to the T cell antigen receptor (TAr) on the effector helper T-cell (eT^H). The signal is triggered by the TAr-target interaction. A similar picture using CD8 would describe the effector cytotoxic T-cell interacting with its target.

In the period that followed 1970, immunologists evinced no interest in the two-signal model (or associative recognition model as I preferred to call it), except for Coutinho and Möller (57) for whom, like us, validly competing concepts are precious. They, therefore, challenged us frontally, and this led to a set of informative exchanges (58–60) in which each of us clarified our positions without being able to convince the other. Thus we learned always to ask ourselves and our opponents what it would take to result in a change of mind. Important for this exchange was the first attempt on our part to tie together two decision functions, the S/NS discrimination and the determination of the class of this response (59, 60).

In 1978, as I described elsewhere (3, foreword), the formalization of the concept of restrictive recognition of antigen by Langman (61) led us (62) to redefine the mechanism of the eT^H-iT/iB interaction as due to two unispecific cells in which the effector T helper (eT^H) initiated and sent Signal [2] via the class II–restricting element on the target iT/iB-cell. We accepted a dual recognitive–single receptor model of the T cell antigen-receptor (TAr) for this formulation, a position I still hold today.

This proposal for the mechanism of Signal [2] is accepted for the iB-cell (with processing) as an eT^H-iB cell-cell interaction leading to induction (T-independent B cells excepted), but totally rejected for the eT^H-iT interaction because of the presently accepted model that requires that a [peptide-MHC] complex be formed and that the TAr react only with that complex. Instead, an APC is interposed between the eT^H- and iT-cell to yield a ménage á trois, eT^H-APC-iT, with resultant induction of the iT-cell. Some immunologists substitute the APC for eT^H in the delivery of Signal [2] (i.e. an APC-iT interaction). While the two-signal model is not disprovable by the detail of mechanism of Signal [2], it is greatly weakened by such a formulation of mechanism because there is no linked relationship to the reference antigen between what the eT^H-cell recognizes and what

the iTH-cell recognizes. For us, this assumption makes unacceptably fuzzy the mechanism of the S/NS discrimination for any peripheral iT mutant with anti-self specificity and that event is unavoidable. However, this is understandably in dispute (64). This is the present state of the discussion, but it has given the two-signal model a place on the list of healthy polemics for the first time (6, 63, 64).

In the period from 1975 to 1985 immunologists considered the TAr to be an immunoglobulin, based on a wealth of data involving the use of anti-idiotypic sera made against immunoglobulin idiotopes. It looked as though my 1971 formulation of IgT was vindicated. The only problem was that our model of the T cell antigen receptor (61, 62) and our knowledge of the properties of the B cell repertoire (87–89) were incompatible with such an interpretation of the findings. Consequently, in 1980 we (65, p. 196) bluntly argued that there is no chance that the Ig-loci would encode the T cell antigen receptor, although in 1978 (61, 62) this point had been made on less direct grounds. The two-signal model began to pay off in its ability to permit correct evaluation of data. Quite clearly no known property of the Ig-loci, and the BAr specificities that they encoded, predicted or explained the behavior of T cells; in particular, they did not explain restrictive recognition of antigen, making the accepted assumption that the TAr was IgT untenable for us. We had changed our minds; our original assumption (54) had to be wrong in spite of its then prevailing popularity. This denial that the T cell antigen receptor was an immunoglobulin met with quite some resistance (66) at the Symposium in Salt Lake City in 1981.

It is surprising to me that the arguments that predicted that TAr could not be an immunoglobulin are ignored when the TAr structure is considered today. The most popular model simply reinvents what is believed to be the combining site of an Ig molecule (67), leaving one to wonder why TAr and BAr had to be separately encoded. Clearly this model (67) will prove to be wrong.

Why Was the "Two-Signal Model" Ignored for so Many Years?

When the "two-signal model" was first formulated in 1968, immunologists were carried away by the data that prompted the assumption that "naked" antigens were tolerigenic only, but that when processed and linked to RNA they became immunogenic only ("superantigen" was the term of the day) (68). This was for me no solution because it put the S/NS discrimination in the hands of a nonspecific processing/linking system, which could not be discriminatory. I really do not know why this idea was so popular or why it went out of fashion; no believer changed his/her mind and, therefore,

once again we learned nothing from it. Many of the arguments used during this period when peptide-RNA was the buzzword are used today in discussing peptide-MHC. It was well-established that peptide derived from protein could be complexed to RNA and that the complex was immunogenic but the "'so what" was missing. This remains another example of a popular idea with experimental support that disappeared without as much as a whimper because it had no conceptual foundation and its disappearance was irrelevant, independent of whether the idea was actually right or wrong.

Then, in 1974, Jerne (69) impassioned the immunological community by proposing what became to be known as the "idiotype network theory" of immune regulation. During the period 1970–1980, I was largely interested in the structure and origin of the repertoire, and consequently I ignored such challenges to the "two-signal model," particularly when they indicated no knowledge of the existence of this competing model and were, in any case, largely smoke and mirrors. In fact, earlier, in 1972, I had invented an "idiotype network" as a Gedanken experiment to show why "instructionist" or template theories were untenable (32, p. 5). However, by 1981, close to one paper every hour was being published on network regulation of the S/NS discrimination via the idiotype, and once again the most elementary of considerations was being ignored. As if caught in a time warp, we were, once again, being mired in the Landsteiner legacy; a transcendental repertoire cannot help but recognize itself, and therefore, an idiotype network is inevitable. True, but it would be nonfunctional, and we had long ago put aside the question of how many angels can dance on the head of a pin. However, I am as much a Jesuit as is Jerne, so let us imagine that a transcendental ("complete") repertoire were functional. In that case, I recall that (i) effector antibody, because it carries out a destructive ridding activity, cannot be used to regulate the S/NS discrimination, and (ii) suppression cannot regulate the S/NS discrimination (3, 6, 51). Further, a good theory must be compatible with some possible mechanism. The humoral immune system cannot be regulated (or integrated) by iB-iB cell-cell interactions via their antigen receptors, membrane Ig; nor can it be regulated by iB-secreted Ig interactions via the combining site of one partner and the V-regions of the other. The first proposal ignores that iB-cells can only receive signals; they cannot send them because iB-cells have no effector function. The second proposal denies the existence of a S/NS discrimination (as does the first proposal) as well as the role of destructive effector functions mediated by secreted Ig, its major function, in fact. Nevertheless, the Jerne idiotype network theory (67) spawned numerous publications on regulation of the S/NS discrimination via suppressor idiotype circuits (70–73), which appeared as a splash and then disappeared

without a ripple. What was it that made the authors of suppressive idiotype circuitry regulating the S/NS discrimination change their minds as to the interpretation of their experiments, or did they? Why were these studies dead-end? Wouldn't we learn to appreciate the role of conceptualization if we knew the answer? Today, the tide of regulation via idiotype networks has receded leaving behind an empty beach; no one has openly changed their mind, and we have no idea what it was that produced the tidal wave or its ebb. If we find ourselves asking 'whatever happened to . . . ?," it is safe to say that we never learned anything from it.

Although from my point of view, the idiotype network theory of regulation lacked logic and rationale, for the sake of history, it is worth this comment. The theory is one more byproduct of the Landsteiner legacy. During the instructionist era the assumption that antigen acted as a template led automatically to a network theory in which antigen molded antibody and the new antibody, acting as a template (experimentally provided in the form of antigen-antibody complexes), induced anti-antibody. This theory, originally proposed by Najjar in 1955, actually permitted him to hail Oudin's discovery of "idiotypes" as the target of the response (74). However, in 1955 the S/NS discrimination was essentially ignored, as was, quite obviously, the obligatory linkage of recognition to effector function. In reviewing the flurry of excitement in the 1980s generated by network theory (69), no fact or credible observation emerged that was unexplainable under or contradicted the "two-signal theory" (3, 75–80), and we were left with two simple arguments of principle.

First, as stated with unadorned clarity by Langman (3, p. 185):

> The central paradox of idiotype network theories is that immunoglobulins must carry out recognition-dependent functions that are nondestructive in order to regulate themselves and they must also carry out recognition-dependent functions that lead to the destruction of the antigens they recognize—whether these antigens are pathogens or other immunoglobulins. There is no way for the same recognitive events to be both destructive (and therefore required to exercise the self-nonself discrimination) and nondestructive (and therefore not required to make the distinction between self and non-self).

Until this point of principle is answered, the many added arguments (75–80) making idiotype networks irrational (independent of whether they are erroneous) seem superfluous. Only then can we learn something from this era of "sound and fury" experimental activity.

Second, idiotype networks require that suppression regulate the S/NS discrimination and, again, as elegantly argued by Langman (49) for the general case, that this cannot be correct.

Finally, I point out that there is no way to prove an irrational theory wrong.

Where Does the Theory of the S/NS Discrimination Stand Today?

A large school of immunologists still believe that suppression is the essence of the S/NS discrimination. I find that position unrewarding until they stop and show how our arguments as to why this is not possible (3, 6, 51) are wrong.

Most immunologists believe that a plethora of different mechanisms contribute to the S/NS discrimination. These "failsafe" pathways are called into play by unspecified mechanisms when the various steps in the basic pathway of the S/NS discrimination break down. I would argue that this position is untenable as there is no way to select for such superimposed "failsafe" pathways because the basic pathway is selected to be sufficiently functional; were this not true, it would be unselectable. After all, when a system of failsafe mechanisms fails, it fails by failing to be failsafe. Autoimmunity exists!

A significant number of immunologists believe that there is an inherently tolerizable-only stage (á la Lederberg) preceding the stage when the iT/iB cell emerges as both tolerizable and inducible. This would reduce autoimmunity to a vanishingly rare case, which is contrary to fact and unselectable (51; see also discussion Ref. 81, pp. 168–170).

Lastly, at the tail end of the distribution there are those immunologists who believe that there is a circumscribed idiotype network that regulates self responses only, leaving the response to nonself to your favorite mechanism. There is no way to put that residue to rest; it is the kind of proposal that time takes care of as it has no unique predictive consequences.

These other views are analyzed elsewhere (3, 4, 6, 75).

The Fifth Era (1980–1993): Coming to Grips with GOD

Actually, it was the problem of the generator of diversity (GOD) that dominated our attention from 1970 on. For us, conceptually, the S/NS discrimination seemed solved except for the question of the origin of the primer, an eT^H-cell, that Langman and I dealt with in the mid-1980s (3, 6, 39, 40, 82). However, the structure and nature of the repertoire was quite ill-defined as a problem and it became pre-occupying.

Were We Really that Smart?

In 1993 it is fashionable to describe the history of our understanding of the origin of the repertoire as "everyone was a little bit right." I have often wondered who originated that description as it has been repeated in so

many reviews and talks. For everyone to have been a little bit right implies that, today, there is a complete lack of awareness that the origin of the repertoire has no correct conceptualization. Thus there is no way to tell which little bit was right. All that the Pollyannish "everyone was a little bit right" shows is that we cannot distinguish between the principle of an argument and the detail of mechanism. No matter who was a little bit right, everyone was a whole lot wrong; and it is from the whole lot wrong that we learn. The major methodological problem is two fold.

First, most of "the little bits that were right" were guesses, not argued or reasoned choices between alternatives and, therefore, often presented in the wrong context. As there are a limited number of solutions to most problems, if a sufficient number of guesses are made someone always comes close to being a little bit right. That is why Nostradamus is considered to be clairvoyant.

Second, most guesses were based on analogy (some quite far-fetched) or on elegance and parsimony, which is a good way to be misled. Few were based on function and its corollary, an evolutionarily selectable pathway; that is the basis for sound biological reasoning (ode to Burnet).

In all of the discussions, most ignored was that the term "molecular biology" has two aspects, molecular and biological, and these must be compatible and reducible to one another. We were overburdened by the Landsteiner legacy that had uncoupled the repertoire from effector function.

There was no adequate theory of the origin of the repertoire (the generator of diversity) in 1970. Such a theory would have had to answer three questions:

1. What does the germline specify and how does evolution maintain it (i.e. the high copy number repertoire)?
2. When (and why) is somatic diversification of the "germline repertoire" required, and how extensive must (or can) it be (i.e. the low copy number repertoire)?
3. What is the contribution of the "germline" encoded repertoire (Stage I or high copy number repertoire) and the "somatically derived" repertoire (Stage II or low copy number repertoire) to the effective response to nonself?

We had to abandon the terms "germline" and "somatically derived" due to their confused use after the discovery of recombination and splicing of gene segments.

There Were Many GODs To Choose From

In 1968, Lennox and I (83) had categorized the views of how the repertoire is generated by the table shown below:

Table 1 Models of origin of varieties in antibodies[a]

Association of genes v and c in germ line	Variety originates in v during: Evolution germ-line models	Differentiation somatic models
v and c form one cistron, vc	Szilard	Brenner & Milstein
v and c are separate cistrons	Dreyer & Bennett	Lennox & Cohn

[a] See ref. 83 for references.

Lennox and I (83) concluded from genetic evidence and evolutionary reasoning that V and C had to be separate gene segments (then referred to incorrectly as cistrons) that were joined somatically as VC to produce the functional transcription unit or cistron. This unit was postulated to undergo somatic mutation to increase the size of the repertoire. It was inconceivable to us that the genome could maintain as many as 10^3 V-gene segments associated with 1 C-gene segment per locus in order to generate by random complementation 10^6 specificities (10^3 $V_L \times 10^3$ V_H) but, admittedly, this was hardly an argument of substance. However, we viewed the potential or total possible repertoire as quite degenerate. This implied to us that the loss of any given specificity could not be felt due to this degeneracy, and the "germline" encoded repertoire would have to shrink until loss had an evolutionarily selectable consequence. This would result in a repertoire of inadequate size encoded in the germline. Only a somatic diversification mechanism to increase its size using this shrunken "germline" repertoire as a substrate would be impervious to degeneracy and, therefore, evolutionarily selectable. This is the reasoning that led us to a two-stage structure for the repertoire; however, this concept was not properly developed until much later because we were totally unprepared to challenge the paradoxes of the Landsteiner legacy derived from the uncoupling of repertoire and effector function.

During this same year (84), I proposed a mechanism for haplotype exclusion based on feedback of the product of gene rearrangement on further rearrangement. If the time taken to successfully rearrange V → C to form a transcription unit were long compared to the time it takes the translated L- or H-chain to shut off further rearrangement, acceptable haplotype exclusion would be accomplished. The doubles under this model would be a function of the ratio of the two times. This model was reinvented in the late 1980s, referred to by us as H-STOP-H and L-STOP-L (4, 25); it turned out to be wrong. The reason that evolution did not seize on that solution is that it could not have regulated L-chain isotype exclusion (κ

and λ). [As an aside, at this point it might have been asked, why are two isotypes advantageous? This question was to be raised again as a byproduct of the mechanism of haplotype exclusion in mice (4, p. 73).] Further, to make joining take long enough compared to shutoff by the product (i.e. $\sim$100:1) might have been unselectable compared to the competing pathway described by the stochastic model (4, 25, 85).

Our first step was to find a way to count *functional* V-gene segments, and our second step was to establish whether somatic mutation diversified the rearranged V-gene segments. We were asking the correct questions but answering them purely empirically, not by a reasoned theory. Nevertheless, both steps were accomplished in 1970.

Two principles were established.

First, the number of germline specified V-gene segments that are expressed as *functional* by plasmacytes (isolated as myelomas) can be counted as being those differing in their framework (FW). Any two amino acid segments differing in FW must be encoded by two different V-gene segments. In 1970 with limited data (admittedly no excuse), I estimated this to be $\sim$20 (86), but by 1973 it was clear that the functional V-gene segments were between 50 and 100 (87–89). I received no end of kidding about the gradual upgrading of my calculation as more data appeared. It was extrapolated that by 1975, I would be at 10^3 V-gene segments per locus indistinguishable from Hood's estimate. This did not happen because there is a difference between the principle of the argument and the sufficiency and accuracy of the data. In fact, the 1973 value of $\sim 10^2$ functional V-gene segments per Igκ- and IgH-haplotype in mouse is close to correct and supported by detailed Protecton calculations (4, 25).

One other comment is worth stressing at this point. As we clearly had in mind, a model of diversification based on two stages, that is, a "germline encoded" repertoire that was "somatically" diversified, a central question was raised: What is the selection pressure maintaining the "germline"? The proposal made in 1970–1974 (86–89) was that the given $V_L V_H$ pairs were selected for the specificities of unique survival value that they encoded, and the 'germline" (STAGE I or high copy number) encoded repertoire resulted from random complementation of these V-gene segments. The selection by carbohydrates on pathogens and by autogenously generated macromolecular waste ("housekeeping" antigens) was justified later (65), by arguing that these substances vary slowly enough to be tracked by the mammalian genome. A viral protein, for example, would mutate to escape recognition too rapidly to be a selective pressure on the genomic V-gene segment. Thus we had the elements of a STAGE I repertoire defined by the mid-1970s, but an essential element was missing, namely, its functional

role as part of the whole and as a way to estimate the limits on its parameters. That came much later (4, 25).

Second, the STAGE I (high copy number) repertoire was somatically varied by mutation to yield the STAGE II (low copy number) repertoire. This was demonstrated unambiguously by Weigert et al (90) in 1970. The logic of this experiment is so elegant it is worth summarizing.

We knew by direct assay, that the κ:λ_1 ratio in murine serum was $\sim$30:1. Because we could not assume uniform antigenic selection on all functional germline V-gene segments, it was argued (90) that there are $\sim$50 V_κ-gene segments and one $V_{\lambda 1}$-gene segment. There had to be a reason why the single $V_{\lambda 1}$-gene segment was maintained in a separate locus, and this reason had to be unique antigenic selection that would disproportionately raise the expressed level of $V_{\lambda 1}C_{\lambda 1}$ thus lowering the observed κ:λ_1 ratio. Eventually, weighing the pluses and the minuses, we settled for an order of magnitude estimate of $\sim 10^2$ functional V_κ-gene segments. How did we know that there was only one $V_{\lambda 1}$-gene segment? The sequences of 10 independently derived λ_1 chains showed 6 to be identical and 4 to vary from them by one or two amino acids (one or two base substitutions) (90). The inevitable conclusion was that there is one $V_{\lambda 1}$ and one $C_{\lambda 1}$ gene segment encoded in the genome and that $V_{\lambda 1}$ is varied by somatic mutation and selected for by antigen. How did we know this latter?

Wu & Kabat (91) had introduced an astute way to analyze the sequence data, referred to as a hypervariability plot. It has been one of the most misinterpreted aspects of the sequence data, although when interpreted correctly, it is extremely useful (65, p. 163). Given a properly balanced somatic mutation theory, complementarity-determining (CD) residues will be found to be hypervariable with respect to framework (FW) residues. However, hypervariability of a given position does not imply that it is CD; it only suggests where to look. There are many reasons that a given position could be hypervariable without being CD. To recall once again the logic, if it rains the fire hydrant will be wet, but if the fire hydrant is wet it does not mean that it has rained. There is no reason to expect discernible hypervariability if the total repertoire is encoded in V-gene segments expressed solely by $V_L V_H$ complementation. However, as Weigert et al (90) showed, the mutational replacements in the sequenced λ_1 chains were all present in regions defined as hypervariable by Wu & Kabat (91). This implied antigenic selection as there is a $\leqslant$20% chance that a mutation would occur in a CD codon compared to a FW codon, and of the 7 base replacements, all were in hypervariable regions (probability without selection $= (0.2)^7 = 10^{-5}$). This striking result told us that what is described as CDR1 and CDR2 by hypervariability contained CD positions

by what we referred to as a pedigree analysis. This analysis is still today the primary criterion for whether a position is CD or FW (see discussion pp. 162–187 in Ref. 65).

In summary, by 1980, we (65) knew that for mouse (and likely human):

1. The "germline" repertoire in the absence of somatic mutation (i.e. the STAGE I or high copy number repertoire) was derived by V-gene segment complementation of 10^2 V_L × 10^2 V_H pairs, of which 10^2 V_LV_H pairs were evolutionarily selected for the specificities of immediate survival value that they provided. Once selected, this high copy number repertoire would be maintained over time if other of the 10^4 derived V_LV_H complements had a special selective advantage.
2. This STAGE I (high copy number) repertoire is the substrate for somatic mutation that yields the STAGE II (low copy number) repertoire.
3. We (65) had rejected all other diversification models, even the one that was, much later, to become the dominant model, referred to by us today as the "neogermline" theory. We argued that the D- and J-gene segments were primarily (evolutionarily selected as) framework, not complementarity-determining, and it was erroneous to calculate a functional repertoire size by multiplying "bits and pieces" to arrive at transcendental numbers (the Landsteiner legacy).

What was missing in 1980 remained the obligatory relationship between recognition and effector function; that relationship was to be established and clarified in 1987 (4, 25).

In order to appreciate this central point, let us consider what were the other views on the generator of diversity during the period 1970–1990?

The Other Views: (1970–1990): Was Everyone a Little Bit Right?

Most immunologists took the position referred to as the "germline theory" that the totality of the repertoire was encoded in the V-gene segments at the IgL- and IgH-loci and revealed by random complementation between their products (92). If the repertoire were guessed to be 10^6, then there must be 10^3 V_L and 10^3 V_H gene segments in the germline. This position seemed simple enough and was attractive because no special mechanisms of diversification needed to be superimposed. However, the statement, "All of the specificities expressed by the individual are encoded in the germline," is an assumption, not a theory. To make it a theory, the selection pressures that maintain the germline must be considered, and this requires an understanding of the relationship between recognition (the repertoire) and an effective effector function. This same comment applies to all of the

"big bang" theories of GOD (see p. 184 in Ref. 65) because they are indistinguishable in consequence from the "germline" theory. It was not the detail of mechanism [recombination, translational or transcriptional scrambling, episomal ("minigene") insertions in CD regions, mutation of a single V-gene segment, etc] that rendered these theories untenable; it was the consequence, namely a vast repertoire with each and every specificity in single copy (i.e. a uniquely low copy number repertoire), that had to be insufficiently functional in protecting the individual. This was not a time when the linkage between recognition and effector function was realized to be an obligatory element that must have had a definable evolutionary pathway.

Even today this same error is perpetuated by deriving the size of the repertoire by multiplying the number of gene segments by junctional diversity and complementation in order to arrive at a transcendental repertoire (i.e. the neogermline theory), in keeping with the Landsteiner legacy. This is an absurdity as such a repertoire would be nonfunctional and is, of course, not selectable. Is there a paper on repertoire or a textbook that does not present the "numbers racket" calculation as the eighth wonder of the world?

The repertoire to be both functional and evolutionarily selectable must be formed in two stages. STAGE I ("germline") had to result from evolutionary selection on the genome. The STAGE II repertoire must result from somatic diversification of the STAGE I repertoire functioning as the substrate. This required evolutionary selection for a diversification mechanism (e.g. hypermutation) not a particular repertoire. The STAGE I repertoire is per force high copy number and the STAGE II repertoire, low copy number, although we did not appreciate these two key characteristics required for function until much later (25).

We were to learn that, in fact, the high copy number repertoire can be read out of the genome by several mechanisms (4, 7):

1. Rearrangement of $\sim 10^2$ V_L- and $\sim 10^2$ V_H-gene segments to yield by expression and random complementation $\sim 10^4$ $V_L V_H$ pairs, the repertoire of high copy number. This is the exchange cassette mechanism used by mouse and human.
2. Rearrangement of 1 V_L- and 1 V_H- recipient gene segment that is varied by controlled gene conversion from $\sim 10^2$ $(V_L + V_H)$ donor gene segments to yield $\sim 10^4$ $V_L V_H$ pairs, the high copy number repertoire. This is the copy cassette mechanism used by some birds, like chicken.
3. Then there is a mixed mechanism (at present not definitively worked out) in which a small number of recipient V_H-gene segments are varied by gene conversion and complemented to a family of rearranged and

expressed V_L containing products, the high copy number repertoire of the rabbit.

4. Lastly, there is the genomic encoding of rearranged VC transcription units, (genes) as found in sharks, the expression of which yields the high copy number repertoire.

These mechanisms have been discussed in forums where the various points of view were aired (4, 7, 25). The varicty of mcchanisms makes the terms STAGE I and II ambiguous (see discussion of Ref. 7), but, for this analysis, the terms are adequate. For our present purpose the mechanism is less important than the principle that evolution had to select for a high copy number repertoire in order to reach the present day stage of a functional system. In no sense could anyone "be a little bit right" if this two stage construct was overlooked. In essence, if everyone were a little bit right, who would be cited as having a correct understanding even today?

Soaring to the Heavens with Icarus

What was the selective pressure for a STAGE I repertoire of high copy number?

Before I give our answer to this question, it is important to consider Jerne's position. He was the only immunologist who seemed to understand the requirement for a two stage unfolding of the repertoire. Accepting that general formulation, Jerne introduced an assumption that cryptically negated the raison d'être for having a two stage rather than a "big bang" construct. In 1970 at the Brook Lodge Symposium Jerne (93) presented us with the following theory:

There are N V_L- and N V_H-gene segments. A total set of N^2 germline specified V_LV_H pairs are selected for their recognition of the major histocompatibility antigens (MHC) of the species. N^2 V_LV_H comprising the STAGE I repertoire was postulated to be divided into two subsets: Subset I is specific for the self-MHC and Subset II, for the nonself or allo-MHC of the species.

Subset I reacting with self-MHC is eliminated by tolerance but mutants of it that no longer recognize self-MHC and are putatively anti-nonself, accumulate to compose the repertoire (STAGE II). Subset II is functionless because an individual never encounters as antigen the allelic products of the MHC of the species.

We know today that this model is wrong for humoral antibody; however, more important is "How rational was Jerne's formulation of the two stage model?"

First, only N given V_LV_H pairs can be selected for their specificities, anti-the-species-alleles-of-MHC, in which case random complements of them, $(N^2\text{-}N)V_LV_H$ pairs, cannot retain this same family of specificities. This should have been particularly obvious as mutants of a given pair were postulated to be anti-non-MHC. When specificity is dependent on a complementing system of V_L- and V_H-gene segments there is no way to select for all of them to be anti-species MHC alleles; only N V_LV_H pairs of the total N^2 can be selected as anti-MHC.

Second, there was no rationale to choosing anti-MHC as the target of the selection; selection for recognition of any set of different self-components, polymorphic or not, would have satisfied Jerne's theory. Of course, Jerne (93) cleverly guessed that the target is anti-MHC because of the Simonsen phenomenon, but that particular target never became an essential part of his theory. I recall that this theory was proposed prior to the discovery of restrictive recognition of antigen by effector T cells. The argument used by Jerne that recognition of self-MHC by B cells anti-self MHC is essential to the viability of the embryo only restates the proposition that this specificity (i.e. anti-MHC) is assumed to be evolutionarily selectable.

Third, the whole point to a two-stage generation of the repertoire is lost if the first stage (high copy number or Subset II) is useless to the functioning of the immune system. The model reduces to the equivalent of all other "big bang" models, as a repertoire of functional mutants uniquely in single copy would be generated.

Fourth, Jerne revealingly asks, "Does the model require allelic exclusion?" He is forced to fall back on the argument that allelic exclusion is simply a fact as, in his way of thinking, this model of diversification does not predict or require it. Without haplotype ("allelic") exclusion each cell would express 4 antibodies (to use a simplified case). If 1% of the germline repertoire were anti-self MHC, then the selectable subset I would be only 3.9% of total, and it would consist of cells expressing [1 anti-self MHC + 3 anti-allo MHC receptors] per cell, whereas subset II, the unselected population would be 96% of total and contain cells with 4 receptors, all anti-allo MHC. The cells with [2 anti-self MHC + 2 anti-allo MHC], [3 anti-self MHC + 1 anti-allo MHC] or 4 anti-self MHC would be lost by tolerance as escape by mutation would be too rare. Subset II is functionless, and the repertoire derived by the mutation of the anti-self MHC to anti-nonself in subset I would flood the animal with the passenger anti-allo MHC upon induction, but so what? As this antibody would be without consequence, Jerne points out that "formally this would be compatible with the selection mechanism" that he proposed; hence, his fall-back position to haplotype exclusion is simply a fact.

I developed Jerne's point to show that his insistence on this argument

reveals that he had no understanding that haplotype exclusion is evolutionarily driven by the requirement for a S/NS discrimination, not for the generation of diversity (the Burnet/Lederberg error); thus, there was no way that he could have explained haplotype exclusion in the context of a model of diversification. It was and still is simply irrelevant. I was always amazed by this argument as Jerne was a leader among those who insisted that "unispecific clonality" was the most important law of immune responsiveness. He voiced this belief most emphatically in the last line of his "complete solution of immunology." In principle, immunology was solved in 1957 when Burnet published his "Clonal Selection Theory of Acquired Immunity" (96). It would be a learning experience to have Jerne explain that complete solution today. In any case, here it was 12 years later, and Jerne was to produce a theory that, in his mind, treated haplotype exclusion as being of marginal import and easily compatible with all cells being doubles or quadruples (93). If he had confronted that seeming internal contradiction, he would have put clonal selection in its proper context and possibly would have paused 17 years later to explain whether idiotype network theory (69) requires unispecific clonality (79, p. 368).

The Jerne theory (93) is that the evolutionary selection pressure is uniquely for STAGE I ("germline") encoded specificities that are anti-self as part of the mechanism for generating a repertoire by "mutational" escape to become anti-nonself. The guess that the self in question is MHC-encoded (by chance, correct for T cells, but incorrect for B cells) or a bursal component (7) is irrelevant for his theory (93). As far as the requirements of the theory are concerned, the particular self-components involved are of no consequence (i.e. provided that the interaction between the STAGE I ("germline encoded") antigen receptors and these self-components have an independent selectable function, immune- or non-immune-related). The best illustration of this is Jerne's midstream switch (with no explanation) from the Ig-loci encoding recognition of MHC-specified restricting elements (93) to the Ig-loci encoding recognition of idiotypes (94) as the self-targets in question. This did not add to the generalized theory; in fact, it only weakened the theory considerably (76).

The assumption that the germline is selected upon uniquely to encode anti-self has left a residue of confusion in the thinking of immunologists that revolves around the meaning of "self." Every time a germline encoded antigen-receptor molecule (i.e. B cell or T cell) is found to have specificity for an autogenous molecule, the Jerne theory (93) is invoked. This is misleading as there is no way to select for recognition of a self-component, if the selection pressure operates to rid the recognizee, an antigen-receptor, by a self-nonself discrimination (76, pgs. 20–25).

It will be argued that I am splitting hairs. After all, every component of

a n animal that participates in a self-nonself discrimination also participates in the physiology of the animal. However, *the evolutionary selection on any given "self" component is based on its role in the physiology of the animal, not on its role in the self-nonself discrimination. The evolutionary selection pressure for a self-nonself discrimination operates on the immune system, not on the self-component.* In the cases where an interaction between membrane Ig (mIg) and some component is essential to the ontogeny from a stem cell to an iB-cell, this component has not been evolutionarily selected to behave in some special way with respect to the S/NS discrimination (e.g. to select a repertoire in the Jernerian sense). Rather, it is important in the differentiation of the system as might be the interaction between a receptor and any hormone or other surface component (7).

I conclude, therefore, that Jerne's theory (93) was simply irrational. Nevertheless, the theory posed some of the right questions and focused thinking on a fundamental one, namely, what maintains the germline? Jerne remained with variations of this theme that the germline is selected for anti-self, in all of his succeeding formulations, although he changed his "self-antigen" from MHC to Ig idiotypes without adequate rationale.

Where Was I During This Period?

As I read my early papers (1968–1970) I can only blush at my lack of perceptive thinking. This was not due to an insufficiency of data but to an inexperience in handling biological complexity. Over time, I was fortunate to be surrounded by colleagues from nonimmunological and nonmedical fields, who looked at the facts of immunology with Cartesian minds and were trained to analyze problems by interactive discussion. These colleagues created a small isolated sanctuary in which new ideas and critical thinking could flourish. As a consequence, this group that included Weigert, Bretscher, Bevan, and Langman produced some of the most important conceptual papers of the past two decades.

Haplotype ("allelic") exclusion as a requirement of the S/NS discrimination using self-indulgent generosity might be said to have its roots in our 1970 paper (48, footnote 40) but it took until 1987 (25) to state clearly why and how the S/NS discrimination drives haplotype exclusion and to define its mechanism and limitation (i.e. Protecton theory). This question then became only one of many elements in understanding the origin of the repertoire.

The first details of a two-stage model were formulated in 1970 (86), and it was to develop in our minds as the competing model to that of Jerne (93). It had the following elements:

1. The STAGE I ("germline") repertoire was selected for specificities of

survival value; one example I used at the time was anti-a(1-3) dextran. Given N V_L- and N V_H-gene segments, N V_LV_H complements were selected for such specificities and $(N^2-N)V_LV_H$ complements were of random or unselected specificity, anti-nonself and anti-self in some ratio, K(4).

2. This STAGE I ("germline") repertoire was the substrate for somatic mutation, which yielded the STAGE II ("somatic") repertoire.

A suggestion for counting functional germline V-gene segments was proposed for the first time, and based on the study of Weigert et al (90) and the analysis of sequence by Wu and Kabat (91), a reasonable two stage somatic mutation model was formulated (86). The thrust of the argument depended on the evolutionary selection by nonself antigen. Of course, I was guilty of ignoring the relationship between recognition and effector function, not to mention, as I already pointed out, my then biased interpretation of the sequence data for which I properly did ample penance and from which I learned a measure of dispassionate thinking in the face of my passionate opinions.

In 1971, Simonsen and I (95) summarized an international workshop on "theories of antibody formation" from which I take this quote:

> "Somatic mutation models do not have difficulties explaining germline evolution because each V gene is selected upon independently. However, the requirements for the initiation of the response to antigen are ill-defined. Estimates must be made of the minimum number of germline genes and mutation frequencies required to give reasonable diversity in the time period for appearance of native responsiveness using reasonable somatic selection mechanisms."

This is what Protecton theory accomplished (4, 25).

In 1972 (32), I tried to derive the consequences for the immune system of two laws:

1. There is no way that a cell can use the configuration of any given molecule to construct directly a protein complementary to it.
2. The self-nonself discrimination cannot be encoded in germline genes.

The first law, a denial of instructionism was interpreted to mean that only selectionist theories are acceptable. The second law implied that for immune systems (not for defense mechanisms of invertebrates and plants) the S/NS discrimination must be learned.

Given these two laws, it became obvious that "clonal selection" was a corollary derivable from these laws, not a primary proposition (32, p. 9). Immunologists were at the time too fired by slogans, not concepts, to appreciate the primary assumptions. Although I clearly saw this at the time, I missed making the simple statement that unispecific clonality

(haplotype exclusion) is driven by the evolutionary selection pressure exerted by the S/NS discrimination. Instead I posed the question and answered it thusly:

> Is there an evolutionary rationalization for the finding that antibody-secreting cells are unispecific? The selective pressure for the specificity of an antibody molecule due to the self-nonself distinction is maximized when one cell expresses one antibody (allelic exclusion operates) (32).
>
> "[T]he selective pressure for allelic exclusion (one cell-one antibody) would be high because: (a) the wasted production of antigen-sensitive cells (with loss of hard-earned specificities) as well as of antibody (upon induction) would be minimized, and (b) the induction of antibody unrelated to the immunogen puts the animal in great danger of autoimmune disease (32)."

The selective pressure for antibodies to be specific is also the S/NS discrimination, and this selection would be severely dulled if cells were oligospecific (see My View of the World). However, it took until 1987, thanks to the critical and integrative thinking of Langman, for these formulations to become precisely definable (4, 25).

Then, too, there existed the problem posed by carrying a significant set of V-gene segments in the STAGE I repertoire.

> Since the self-nonself discrimination cannot be germline encoded (second law) each individual must under a germline model carry all of the structural V genes coding for both self- and nonself-recognition. These genes must be fixed in the germline because the specificities they code for are of selective advantage at the level of the individual. The structural V genes coding for nonself are understandably of selective value but those coding for self are either silent (tolerance) or deleterious (autoimmunity). Consequently, antiself specificities cannot be fixed in the germline by antigenic selection. The way to resolve this paradox posed by the second law is to assume that it is impossible to construct a system consisting of a large number of complementing nonidentical subunits in which the fixation of V genes because they confer useful anti-nonself specificities does not simultaneously also fix anti-self specificities (32).

Thus the elements of a two stage theory were beginning to take shape. As an aside, today, I would view the immune system in terms of the three Langman laws (3) as they focus sharply on the physiological imperatives.

From 1972–1974 the two stage theory of the origin of the repertoire was tested against experiment and found to hold its own remarkably well (87–89). We relaxed and turned to the analysis of the associative recognition model ("two-signal theory") in terms of the then emerging body of experiment testing its validity. This, too, held its own remarkably well (58,

59). The general ignoring of the existence of this model by the idiotype networkers has always intrigued me and, of course, by not considering competing theories, resulted in the failure of idiotype network theory to go anywhere (75–80). On the one hand, proponents of idiotype networks touted "unispecific clonality" as the most important and fundamental idea of immune function (96), while, on the other hand, they bulldogged a theory that was impervious to whether a cell produced one or two or more antibodies (69). Validly competing theories are precious and should always be carefully nurtured and evaluated side by side.

By 1980, our passion had become "restrictive recognition of antigen" a subject I cannot analyze here. However, its impact on what I have discussed was enormous, driving us to rethink all aspects of the detail of mechanism. Further, the role of interleukins was uncovered, and we needed to integrate these findings. A good theory always lends itself to encompassing new findings. I can only invite the reader to consider those extensions (3, 6, 8, 40, 82).

The Protecton Is Born

The unification of the theoretical "bits and pieces," two-stage repertoire, haplotype exclusion, S/NS discrimination, Ig structure and signalling, etc, was to appear as Protecton theory, the first comprehensive and integrated look at the humoral immune system (4, 7, 25, 97, 99). All that was needed was a way to connect recognitive capability to effector function. Langman and I (4, 25) did this by invoking three assumptions:

1. Ig antibodies carry out their effector functions in a concentration-dependent manner.
2. All of the parameters determining Ig effector function can be summed as a threshold concentration required to form sufficient antigen-antibody complexes in order to ensure ridding of antigen [the Talmage (38) proposition].
3. The time taken to reach the threshold concentration must be short enough to provide protection before the growing pathogen reaches a lethal level.

These assumptions lead to the conclusion that the humoral immune system is modular in construct. This is the new essential concept. This module, referred to as a *Protecton*, is the smallest sample that can be taken from the B cell population of an animal and still retain all of the functional (protective) properties of the whole. As the size of an animal increases so does the number of Protectons. For example, a hummingbird consists of 1 Protecton, a mouse of 10 Protectons, a human of 10^5 Protectons, and an elephant of 10^7 Protectons. All Protectons are functionally equivalent,

a conclusion of profound importance. For its primary response, an animal with a single Protecton like a hummingbird is protected at the 90% level by its humoral immune system as is an elephant of 10^7 Protectons. As a consequence of *Assumption* 1, animals are protected per ml, not per animal. Having more than one Protecton offers little increase in the effective level of primary protection per animal; having less than one Protecton leads to a sharp decline in the level of primary protection. For example, tadpoles with a tenth of a Protecton have roughly a 10% chance of surviving a random infection. However, an adult frog has a 90% chance of survival, the same as that of an elephant. The solution that evolution has taken in this case is to produce and waste many tadpoles to arrive at one frog.

Consider now a Gedanken experiment. A mouse-sized sample of the B cell population of the elephant will protect a mouse. If the mouse-sized sample of B cells from the elephant can protect the mouse, then the B cell population of the elephant must be repetitive. If we extrapolate to the smallest size sample of B cells that will remain adequately protective (a hummingbird's worth), then we have a Protecton. It is the characteristics of this unit, a Protecton, that evolution defined the humoral immune system by making it the unit of selection.

Secreted Ig functions as the effector mechanism. This mechanism is dependent on the concentration of antigen relative to the affinity of the combining site and on the concentration of aggregated bound Ig relative to its affinity for the effector mechanisms (C′ lysis, phagocytosis, ADCC, etc), which rid Ag bound to Ig. All of these factors can be summed as an average threshold concentration of antibody minimally required to activate antigen removal systems (*Assumption* 2). This amount of antibody (estimated to be around 10ng/ml) dictates how many iB-cells must respond to an antigen in order that a protective level of antibody be produced in a short enough time (estimated to be $\sim$1 week); this is *Assumption* 3.

What are the characteristics of the minimum iterated unit of humoral protection (i.e. the Protecton)? This is where the surprises emerge. Each Protecton must respond rapidly enough to rid the pathogen, and the response must be specific for the pathogen (i.e. the response to the pathogen must not entrain a threshold level of anti-self that would trigger autoimmunity). This specificity requirement is what necessitates haplotype exclusion. I leave the reader to evaluate Protecton theory as it has been developed elsewhere (4, 7, 25, 99).

The conclusions dependent on derivations from the three assumptions (4) provide a *totally* competing view of the humoral immune system. At no time since the 1950s, when "instructionism" versus "selectionism" was polemic, has such a clear-cut dichotomous situation been placed before immunologists for consideration.

Consider the following partial list of dichotomous conclusions derived for mouse from:

Protection theory	Majority opinion
1. Double mIg^+ B-cells are generated at the 5% level.	1. No ($\lesssim 1\%$) double mIg^+ B-cells are produced.
2. The D-N sequence variability is used to adjust the level of haplotype exclusion (i.e. D-disaster obtains).	2. The D-N sequence variability translates directly into functional antigen-binding specificity (i.e., D-diversity obtains).
3. In mouse around 90% of B-cells are "non-functional."	3. In mouse, all B-cells are "functional."
4. Signal [1] to the iB-cell is conformationally driven.	4. Signal [1] to the iB-cell is aggregationally driven.
5. The functional available repertoire is $\sim 10^5$. "Completeness" is nonsense.	5. The functional available repertoire is $> 10^{10}$. "Completeness" is fundamental. (The Landsteiner legacy.)
6. The half-life of serum Ig is long (>90 days).	6. The half-life of serum Ig is short (<30 days).
7. In mouse, the antigen unselected κ:λ ratio is around 1.	7. In mouse, the antigen unselected κ:λ ratio is around 10–20.
8. Affinity maturation is a second order phenomenon for function.	8. Affinity maturation is the raison d'être for somatic mutation (i.e. "fine tuning").
9. Regulation by idiotype networks is ruled out.	9. Regulation by idiotype networks is central.
10. Only $\sim 10^2$ self-antigens are important targets for the S/NS discrimination by the humoral immune system.	10. Every gene product is a self-antigen acting as an important target for the S/NS discrimination by the humoral immune system.

I make the following brief comments on these conclusions.

1. Haplotype exclusion cannot be perfect. If no doubles existed there would be nothing to select upon. The boundary condition maintained by evolution is to make autoimmunity not limiting as a factor in survival.
2. D-diversity leads to the modern neogermline or "big bang" theory of the repertoire. No arguments have changed, nor have problems been solved. D-disaster is evolutionarily selectable; D-diversity is not. In essence, D_H was selected not for a role in determining the repertoire but for a function dependent on its being framework (i.e. D-disaster). We suggest, as most likely, haplotype exclusion, but other roles are possible (4, 7) and under discussion.
3. So ingrained is the assumption that all murine B cells are functional, that it became the basis for Nossal's demonstration of B cell anergy.

Mice responsive or rendered unresponsive to a given antigenic determinant have the same number of antigen-binding B cells; ergo, anergy. However, if 90% were not tolerizable (nonfunctional), then the regimen leading to unresponsiveness could only reduce the number of antigen-binding cells by 10% (i.e. not detectable) (81).

In chicken, where the repertoire becomes self-renewing, all iB-cells are functional (7). Therefore, this experiment should be repeated in chickens.

4. "Has immunoglobulin come to a sticky end? (5)" Read it and refute the arguments. Answer the question, "What will it take to change your mind?" and, then, let us have a discussion.
5. The Landsteiner legacy is put sharply into focus. The implied degeneracy, $>10^5$ fold, is not selectable (4, 5) because a transcendental ("complete") repertoire is nonfunctional.
6. Jerne also made this key point (94, pp. 14–17) from another set of data. Upon encountering a pathogen for the second time, the animal is more importantly protected by its serum antibody than its memory B cells.
7. The molecular biology and the cellular immunology are in contradiction. There would be no way to know this without a theory (see discussion in Ref. 98). There is a far reaching question posed. Why have different L-chain isotypes, κ and λ evolved? Why is the $V_{\lambda 1}$ of mouse disproportionately expressed? (4, p. 73)
8. I am using "affinity maturation" to mean that, during certain immune responses, within a clone, higher and higher affinity mutants are selected that account for the overall increase in average affinity of the secreted antibody. It is informative to point out that affinity maturation must be non-existent in birds where the repertoire becomes self-renewing (discussed in Ref. 7). Clearly, if an animal cannot survive a primary infection, a secondary response is of no significance.
9. The necessity to link recognition to effector function cannot be ignored. Positive and negative regulation cannot be mediated by a molecule that, when regulating, carries out a destructive ridding function.
10. The principle of equivalence of Protectons puts limits upon what is effectively a selection pressure exerted by a self component (4).

Today, the conclusions derived from Protecton theory stand in sharp contrast to the conventional wisdom. If Protecton theory is not ignored like the theories of a "two-stage repertoire" and of the "two-signal S/NS discrimination" something of great value will emerge for all of us.

Where To From Here?

In the meantime, rather than looking at the past, I intend to turn my attention to the future by analyzing two problems of immunology, one almost untouched, the other suffering from overkill.

The largely untouched problem is, "how is the class of the response determined?" This question has occupied our thinking over the years (3, 40, 56, 59), but a generalized heuristic concept has not emerged; however, there is now a rich body of experimental data to be analyzed.

The treatment of a problem by overkill is a sure sign that no satisfyingly unified concept is emerging. This problem is that of "restrictive recognition of antigen." While we had begun to address that question by developing a dual recognitive–single receptor model of the T cell antigen receptor, it has been treated as ruled out by immunologists because of their conviction that restricted T cells can only recognize peptides bound in the groove of MHC encoded restricting elements. This whole subject needs reanalysis rather than bandaid solutions dependent on reinventing another dual recognitive construct only marginally explicative of the data (67). It is time for an integrated approach beginning by defining the unit of cell-mediated function, that is, a T cell Protecton (i.e. the linking of recognition to effector function).

Given a good conceptual foundation, the totality can then be gathered into a general theory of great predictability that will allow a new kind of experimentation using the computer to guide and complement that with test tubes. Such a program will permit us to bite off large chunks of the phenomenology of the immune system, thereby reducing the number of competing concepts to a minimum.

This transition from the rear view mirror to the road ahead is the cue that "makes me end, where I began."

Acknowledgments

I dedicate this essay to the memory of Maurice Landy. In the last years of his life he was part of my laboratory, giving selflessly of his wisdom and wealth of experience to all of us, but above all to the younger scientists. Maurice enjoyed reading, criticizing, and evaluating our papers; he helped get the most creative and original of them past the, all too often, hostile reviewers. He taught the innocent among us, grantsmanship, and the more jaded, he taught tenacity. His enthusiasm was unbounded to make us put our thoughts into print where their heat resistance could be assayed. Maurice felt that reflections on experience and history were of major value. Would that he could have read and criticized this essay.

I can only express my indebtedness to Dr. Rodney Langman by making

him share responsibility for any egregious errors of omission or commission. Supported by NIH Grant RR 07716.

Literature Cited

1. Cohn M. 1967. Reflections on a discussion with Karl Popper: The molecular biology of expectation. *Bull. All-India Inst. Med. Sci.* 1: 8–16
2. Mitchison NA. 1990. Better to confess ignorance: In answer to Mel Cohn. In *Cell to Cell Interaction*, ed. MM Burger, B Sordat, RM Zinkernagel, pp. 232–234. Basel: Karger
3. Langman RE. 1989. *The Immune System*. San Diego: Academic
4. Cohn M, Langman RE. 1990. The Protecton: The evolutionarily selected unit of humoral immunity. *Immunol. Rev.* 115: 1–131
5. Langman RE, Cohn M. 1991. Has immunoglobulin come to a sticky end? *Scand. J. Immunol.* 33: 99–109
6. Cohn M. 1992. The self-nonself discrimination: Reconstructing a cabbage from a sauerkraut. *Res. Immunol.* 143: 323–34
7. Langman RE, Cohn M. 1993. A theory of the ontogeny of the chicken humoral immune system: The consequences of diversification by gene hyperconversion and its extension to rabbit. *Res. Immunol.* In press
8. Langman RE, Cohn M. 1992. What is the selective pressure that maintains the gene loci encoding the antigen-receptors of T- and B-cells? A hypothesis. *Immunol. Cell Biol.* 70: 397–404
9. Cohn M. 1982. An introduction to "Le Pap." *Cell. Immunol.* 66: 11–16
10. Luria SE, Delbruck M. 1943. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* 28: 491–511
11. Cohn M. 1970. Anticipatory mechanisms of individuals. In *Control Processes in Multicellular Organisms*, ed. GEW Wolstenholme, J Knight, JA Churchill, pp. 255–303. London: CIBA Found. Symp.
12. Pauling L. 1940. A theory of the structure and process of formation of antibodies. *J. Am. Chem. Soc.* 62: 2643–657
13. Haurowitz F. 1952. The mechanism of the immunological response. In *Biological Reviews of the Cambridge Philosophical Society*, ed. H. Munro Fox, pp. 247–80
14. Haurowitz F. 1953. The immunological response. *Annu. Rev. Microbiol.* VII: 389–414
15. Haurowitz F. 1953. Theories of antibody formation. In *The Nature and Significance of the Antibody Response*, ed. AM Pappenheimer, Jr., New York: Columbia Univ. Press
16. Haurowitz F. 1960. Immunochemistry. *Annu. Rev. Biochem.* 29: 609–34
17. Burnet FM. 1959. *The Clonal Selection Theory of Acquired Immunity*. Nashville, Tenn: Vanderbilt Univ. Press. 209 pp.
18. Landsteiner K. 1945. *The Specificity of Serological Reactions*. New York: Dover. 330 pp.
19. Burnet FM, Fenner F. 1948. Genetics and immunology. *Heredity* 2: 289–324
20. Talmage, DW. 1957. Allergy and immunology. *Annu. Rev. Med.* 8: 240–56
21. Burnet FM, Fenner F. 1949. *The Production of Antibodies*. London: Macmillan. 142 pp.
22. Burnet FM. 1956. *Enzyme Antigen and Virus: A Study of Macromolecular Pattern in Action*. Cambridge: Cambridge Univ. Press. 193 pp.
23. Cohn M. 1978. In Memoriam to Jacques Monod, In *The Operon*, ed. JH Miller, WS Reznikoff, pp. 1–9. New York: Cold Spr. Harbor Lab.
24. Hogness DS, Cohn M, Monod J. 1955. Studies on the induced synthesis of β-galactosidase in *Escherichia coli*: The kinetics and mechanism of sulfur incorporation. *Biochim. Biophys. Acta* 16: 99–116
25. Langman RE, Cohn M. 1987. The E-T (elephant-tadpole) paradox necessitates the concept of a unit of B-cell function: The Protecton. *Mol. Immunol.* 24: 675–97
26. Attardi G, Cohn M, Horibata K, Lennox ES. 1959. Symposium on the biology of cells modified by viruses or

antigens. II. On the analysis of antibody synthesis at the cellular level. *Bact. Revs.* 23: 213–23
27. Attardi G, Cohn M, Horibata K, Lennox ES. 1964. Antibody formation by rabbit lymph node cells. I. Single cell responses to several antigens. *J. Immunol.* 92: 335–45
28. Attardi G, Cohn M, Horibata K, Lennox ES. 1964. Antibody formation by rabbit lymph node cells. II. Further observations on the behavior of single antibody-producing cells with respect to their synthetic capacity and morphology. *J. Immunol.* 92: 346–55
29. Attardi G, Cohn M, Horibata K, Lennox ES. 1964. Antibody formation by rabbit lymph node cells. III. The controls for microdrop and micropipet experiments. *J. Immunol.* 92: 356–71
30. Attardi G, Cohn M, Horibata K, Lennox ES. 1964. Antibody formation for rabbit lymph node cells. IV. The detailed methods for measuring antibody synthesis by individual cells, the kinetics of antibody formation by rabbits and the properties of cell suspensions. *J. Immunol.* 92: 372–90
31. Attardi G, Cohn M, Horibata K, Lennox ES. 1964. Antibody formation for rabbit lymph node cells. V. Cellular heterogeneity in the production of antibody to T5. *J. Immunol.* 93: 94–95
32. Cohn M. 1972. Immunology: What are the rules of the game? *Cell. Immunol.* 1–20
33. Jerne N. 1955. The natural-selection theory of antibody formation. *Publ. Natl. Acad. Sci. USA* 41: 849–57
34. Burnet FM. 1957. A modification of Jerne's theory of antibody production using the concept of clonal selection. *Aust. J. Sci.* 20: 67–69
35. Newcombe HB. 1949. Origin of bacterial variants. *Nature* 164: 150–51
36. Lederberg J, Lederberg EM. 1952. Replica plating and indirect selection of bacterial mutants. *J. Bacteriol.* 63: 409–6
37. Lederberg, J. 1959. Genes and antibodies. *Science* 129: 1649–53
38. Talmage DW. 1959. Immunological specificity. *Science* 129: 1643–48
39. Cohn M. 1983. Antibody diversity 1983: Some elementary considerations, In *Progress in Immunology*, ed. Y Yamamura, T Tada, pp. 840–51. Orlando, Fla: Academic
40. Cohn M. 1989. The a priori principles which govern immune responsiveness, In *Cellular Basis of Immune Modulation*, pp. 11–44. New York: Alan R. Liss
41. Talmage DW. 1986. The acceptance and rejection of immunological concepts. *Annu. Rev. Immunol.* 4: 1–11
42. Cohn M. 1956. On the inhibition by glucose of the induced synthesis of β-galactosidase in *Escherichia coli*. In *Enzymes: Units of Biological Structure and Function*, ed. OH Gaebler, pp. 41–48, 96–98. New York: Academic
43. Mazumdar PH, ed. 1989. *Immunology 1930–1980: Essays on the History of Immunology*. Toronto: Wall & Thompson. 307 pp.
44. Sexton C. 1991. *The Seeds of Time: The Life of Sir Macfarlane Burnet*. Australia: Oxford Univ. Press. 301 pp.
45. Owen RD. 1945. Immunogenetic consequences of vascular anastomoses between bovine twins. *Science* 102: 400–1
46. Forsdyke DR. 1968. The liquid scintillation counter as an analogy for the distinction between "self" and "not-self" in immunological systems. *Lancet.* pp. 281–83
47. Bretscher PA, Cohn M. 1968. Minimal model for the mechanism of antibody induction and paralysis by antigen. *Nature* 220: 444–48
48. Bretscher PA, Cohn M. 1970. A theory of self-nonself discrimination. *Science* 169: 1042–49
49. Langman RE. 1987. The self-nonself discrimination is not regulated by suppression. *Cell Immunol.* 108: 214–19
50. Bretscher P. 1972. The control of humoral and associative antibody synthesis. *Trans. Rev.* 11: 217–67
51. Cohn M. 1987. *The Tolerance Workshop. Proceedings of the EMBO Workshop on Tolerance, Basel*, pp. 3–35. Basle: Editiones (Roche)
52. Landy M, Braun W, eds. 1969. *Immunological Tolerance*. New York: Academic. 352 pp.
53. Gershon RK, Kondo K. 1970. Cell interactions in the induction of tolerance. *Immunology* 18: 723–37
54. Cohn M. 1971. The take home lesson—1971. In *Immunoglobulins*, ed. S Kochwa, HG Kunkel, 190: 529–84. New York: Ann. NY Acad. Sci.
55. Bretscher P. 1973. A model for generalised autoimmunity. *Cell. Immunol.* 6: 1–11
56. Bretscher P. 1974. Hypothesis: On the control between cell-mediated, IgM and IgG immunity. *Cell. Immunol.* 13: 171–95
57. Coutinho A, Möller G. 1974. Immune activation of B-cells: evidence for "one nonspecific triggering signal" not delivered by the Ig receptors. *Scand. J. Immunol.* 3: 133–46
58. Cohn M, Blomberg B. 1975. The self-nonself discrimination: a one- or two-

signal mechanism? *Scand. J. Immunol.* 4: 1–24
59. Cohn M. 1975. The logic of cell interactions in determining immune responsiveness. In *Molecular Approaches to Immunology*, ed. EE Smith, DW Ribbons, pp. 79–107. New York: Academic
60. Cohn M. 1976. Summary discussion of directions in immunobiology. In *Mitogens in Immunobiology*, ed. JJ Oppenheim, DL Rosenstreich, pp. 663–702. New York: Academic
61. Langman RE. 1978. Cell-mediated immunity and the major histocompatibility complex. In *Rev. Physiol. Biochem. Pharmacol.* 81: 1–37. New York: Springer-Verlag
62. Cohn M, Epstein R. 1978. T-cell inhibition of humoral responsiveness. II. Theory on the role of restrictive recognition in immune regulation. *Cell. Immunol.* 40: 125–53
63. Cohn M. 1990. Jottings on Mitchison's I-See-What-I-Believe Immune System. In *Cell to Cell Interaction*, ed. MM Burger, B Sordat, RM Zinkernagel, pp. 215–231. Basel: Karger
64. Langman RE, Cohn M. 1993. A reply to Fuchs's comments on two signal models of lymphocyte activation. *Immunol. Today* 14: 235–37
65. Cohn M, Langman RE, Geckeler W. 1980. Diversity 1980. In *Immunology 1980: Progress in Immunology IV* (4th Intl. Congr. of Immunol.), ed. M Fougereau, J Dausset, pp. 153–201. New York: Academic
66. Cohn M. 1981. Pinch hitting for Hans Wigzell. In *Immunoglobulin Idiotypes and Their Expression*, pp. 831–49. New York: Academic
67. Davis MM, Bjorkman PJ. 1988. T-cell antigen receptor genes and T-cell recognition. *Nature* 334: 395–402
68. Plescia OS, Braun W, eds. 1968. *Nucleic Acids in Immunology*. New York: Springer-Verlag. 724 pp.
69. Jerne NK. 1974. Towards a network theory of the immune system. *Ann. Immunol. (Inst. Pasteur)* 125C: 373–89
70. Dorf MF, Benacerraf B. 1984. Suppressor cells and immunoregulation. *Annu. Rev. Immunol.* 2: 127–58.
71. Green DR, Flood PM, Gershon RK. 1983. Immunoregulatory T-cell pathways. *Annu. Rev. Immunol.* 1: 439–63
72. Germain RN, Benacerraf B. 1981. A single major pathway of T-lymphocyte interactions in antigen-specific immune suppression. *Scand. J. Immunol.* 13: 1–10
73. Dorf MF, Okuda K, Minami M. 1981. Dissection of a suppressor cell cascade. *Curr. Top. Microbiol. Immunol.* 100: 61–7
74. Najjar VA, Robinson JP. 1958. A unified concept in immunologic reactions and allergic states. In *Immunity and Virus Infection*, ed. V.A. Najjar, pp. 71–91. New York: Wiley
75. Cohn M. 1981. Conversations with Niels Kaj Jerne on immune regulation: Associative versus network recognition. *Cell. Immunol.* 61: 425–36
76. Cohn M. 1985. Diversity in the immune system: "Preconceived ideas" or ideas preconceived. *Biochimie* 67: 9–27
77. Cohn M. 1986. The concept of functional idiotype network for immune regulation mocks all and comforts none. *Ann. Inst. Pasteur/Immunol.* 137C: 67–76
78. Cohn M. 1986. Comments for 10th Forum on Immunology, "Is the immune system a functional idiotypic network?" *Ann. Inst. Pasteur/Immunol.* 137C: 173–88
79. Cohn M. 1986. Idiotype network views of immune regulation: For whom the bell tolls. In *Idiotypy & Medicine*, ed. JD Capra, pp. 321–99. New York: Academic
80. Langman RE, Cohn M. 1986. The "complete" idiotype network is an absurd immune system. *Immunol. Today* 7: 100–01
81. Möller G. ed. 1990. Concepts in immunology. *Immunol. Rev.* 115
82. Langman RE. 1992. Was everyone a "little-bit-right" after all? *Res. Immunol.* 142: 316–22
83. Lennox ES, Cohn M. 1967. Immunoglobulins. *Annu. Rev. Biochem.* 36: 365–406
84. Cohn M. 1968. What can *Escherichia coli* and the plasmacytoma contribute to understanding differentiation and immunology? In *Symposium of the International Society for Cell Biology*, ed. KB Warren, pp. 1–28. New York: Academic
85. Claverie J-M, Langman R. 1984. Models for the rearrangements of immunoglobulin genes: A computer view. *TIBS* 9: 293–96
86. Cohn M. 1970. Selection under a somatic model. *Cell. Immunol.* 1: 461–67
87. Cohn M. 1973. Antibody diversification: the somatic mutation model revisited. In *The Biochemistry of Gene Expression in Higher Organisms*, ed. JK Pollak, JW Lee, pp. 574–92. Dordrecht, Holland: Reidel
88. Cohn M. 1974. A rationale for ordering the data on antibody diversification. In

Proc. 2nd Intl. Congr. of Immunol., Brighton, ed. L Brent, J Holborow, pp. 261–84. Amsterdam, Holland: North Holland
89. Cohn M, Blomberg B, Geckeler W, Raschke W, Riblet R, Weigert M. 1974. First order considerations in analyzing the generator of diversity. In *The Immune System: Genes, Receptors, Signals*, ed. EE Sercarz, AR Williamson, CF Fox, pp. 89–117. New York: Academic
90. Weigert MG, Cesari IM, Yonkovich SJ, Cohn M. 1970. Variability in the lambda light chain sequences of mouse antibody. *Nature* 228: 1045–47
91. Wu TT, Kabat E. 1970. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J. Exp. Med.* 132: 211–50
92. Dreyer WJ, Bennett JC. 1965. The molecular basis of antibody formation: Paradox. *Proc. Natl. Acad. Sci. USA* 54: 864–69
93. Jerne NK. 1970. Generation of antibody diversity and self tolerance—a new theory. In *Immune Surveillance*, M Landy, RT Smith, pp. 343–436. San Diego: Academic
94. Jerne NK. 1984. Idiotypic networks and other preconceived ideas. *Immunol. Rev.* 79: 5–24
95. Cohn M, Simonsen M. 1971. Theories of antibody formation. In *Progress in Immunology*, ed. B Amos, pp. 1403–05. New York: Academic
96. Jerne N. 1969. The complete solution of immunology. *Aust. Ann. Med.* 4: 345–48
97. Langman RE. 1992. Molecular economy and antibody function: the evolution of a Protecton. *Int. J. Clin. Lab. Res.* 22: 63–8
98. Langman RE, Cohn, M. eds. 1992. What determines the κ/λ ratio? 47th Forum in Immunology. *Res. Immunol.* 143: 803–39
99. Langman RE, Cohn M. 1992. What is the selective pressure that maintains the gene loci encoding the antigen-receptors of T- and B-cells? A hypothesis. *Immunol. Cell. Biol.* 70: 397–404