Mathematical Challenges to the Neo-Darwinian Interpretation of Evolution
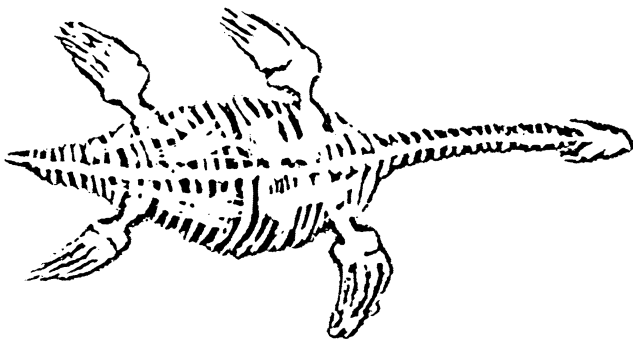
Edited by
PAUL S. MOORHEAD
MARTIN M. KAPLAN

Algorithms and the Neo-Darwinian Theory of Evolution

DR. MARCEL P. SCHÜTZENBERGER

University of Paris, France

DR. MARCEL SCHÜTZENBERGER: Our thesis is that neo-Darwinism cannot explain the main phenomena of evolution on the basis of standard physico-chemistry. Here we stress two points. First that the physical concepts used by biologists are generally more classical (or less imaginative) than the ones occurring in such domains as, say, cosmology. Second that we are not trying to smuggle in extra scientific principles. Thus if we claim that radically new principles are needed we also believe that these have to be found within physics. The nature of the inability of biology to provide a coherent explanation of evolution is best seen when contrasting it with geology. It is certain that no one can work out mathematically every detail of the geological history of the earth. However, for each of the most important phenomena, there exists a simplified model which accounts for it, without mysterious forces, and there is no doubt that this chain of models could be refined *ad infinitum* without gaps and without requiring the verbal argument so often met here that new qualitative effects arise because of the enormous number of small quantitative variations. At no point does geology need to use such phrases as "creation of information", "increase of efficiency", "self-organization", and the like. (My examples are chosen so as to offend no one here, I hope). I intend to restrict my argument to show the existence of a serious gap in the current theory of evolution. The next question (which I will not discuss here) would be to ask how much random mutation and selection would be needed once this gap is filled.

My colleagues this morning have been doing their share of sand reckoning in the manner of Archimedes. From their talks it is clear that even on the most schematic models the number of cycles involved is truly enormous. Thus, when we reach the level of 10^{1000}, whether or not we take a few square roots makes little difference in this cosmos. A second point to which I would like to draw your attention is the fact that nowadays computers are operating within a range which is not entirely incommensurate with that dealt with in actual evolution theories. If a species breeds once a year, the number of cycles in a million years is about the same as that which one would obtain in a ten day computation which iterates a program whose duration is a hundredth of a second. Our ability to play with iteration of this magnitude is quite a new thing, and we can begin to develop some concrete experience with this type of process. It was not so in the time of Fisher and *mon bon Maître* Haldane, and now we have less excuse for explaining away difficulties by invoking the unobservable effect of astronomical numbers of small variations.

To present my argument I need to introduce a schematization of current ideas based on the introduction of three spaces, each endowed with a specific net of proximity relations, or as I shall say for short, a topology—if you forgive my using mathematical jargon.

According to the "dogma" of molecular biology the first level we start with is, ideally, something like a big book written in an alphabet of 20 odd letters. This is the blueprint of an individual, a genotype. Further we have a genic pool, i.e., a collection of such books which are variants of each other. For many clusters of species this collection is not much bigger than the Widener Library; for others, it is at most millions or billions of times larger. I shall

take those books as the elements of the first space, and, since we are not in any way Lamarckian, we have to admit that the proximity relations in this space are of a strictly typographic nature: omission, addition, duplication, transposition, or change of letters, pages, or of chapters—but irrespective of context, or if you allow me, of meaning. This topology, insofar as molecular biology is concerned, is of the same nature as the one which would represent the relations between several copies of the same manuscript typed and bound by a very careful assistant totally ignorant of the language in which it was written. Typically, in the typographic topology, two editions of the same textbook of botany differing only in the fact that one contains the common name of species wherever the other has the Latin name, would be further apart than two editions differing by replacement (or deletion, or duplication . . . etc . . .) by another word or jumble of letters perpetrated in a systematically random manner, of one or two words in each page.

At the opposite end we have the individuals who react to the environment in accordance with their being physico-chemical systems with a given size and configuration. Admittedly, it may be hard to give an abstract formulation of the fact that two trees (or two winged animals, or two protozoa) are "closer" to each other in the topology of phenotypes than are, say, a bush and a bird. However, this system of closeness relations is the one on which we base most of our taxonomy and physiology. It is with reference to this topology that one tries to account for the similarity of the selective effects of the milieu when discussing phenomena of convergence.

In the middle, neo-Darwinism introduces a third space consisting of vectors (i.e. finite sets of numerical parameters) with its usual topology. The coordinates of these vectors are such things as mutation rates, coefficients of viability, etc. Because this is a theoretical object, it is not dramatically surprising that one can predict or simulate Darwinian effects within it. One might question the adequacy of using the parameter space as a model for the phenotypic space and the validity of the reasonings based on it be-

cause in almost every case, both the parameters and the relations between them are strictly hypothetical constructs for which no conceivable direct or indirect cross measure exists. We shall not do it here because we believe that the crucial difficulty is not in relating this theoretical parameter space to the real phenotypic space but in providing a link, however tenuous, between either of them and the space of the chains of amino acids (or the space of genic pools, it does not make much difference) endowed with its specific typographic topology.

Indeed, what we have at each of the two extremes is not even chaos out of which one might believe that a certain regularity could emerge, as it may do in thermodynamic processes, but two systems having structures (topologies) which *a priori* are not more in agreement than in conflict. Now some modicum of agreement is needed if one wants the selection pressure to have the nice effects we are told it has. Otherwise, there is no reasonable probability (say more than 10^{-1000}) that variations in the milieu operate without the genotype having entered a *cul-de-sac* out of which no evolution is possible.

I apologize for being so assertive but here is the point where experience with computers (more seriously, of course, a few mathematical results) comes in. According to molecular biology, we have a space of objects (genotypes) endowed with nothing more than typographic topology. These objects correspond (by individual development) with the members of a second space having another topology (that of concrete physico-chemical systems in the real world). Neo-Darwinism asserts that it is conceivable that without anything further, selection based upon the structure of the second space brings a statistically adapted drift when random changes are performed in the first space in accordance with its own structure.

We believe that it is not conceivable. In fact if we try to simulate such a situation by making changes randomly at the typographic level (by letters or by blocks, the size of the unit does not really matter), on computer programs we find that we have no chance (i.e. less than $1/10^{1000}$) even to see

what the modified program would compute: it just jams. We can specify what it would take to have the random modifications introduced so that a sizable fraction of all programs start working: It is a self-correcting mechanism which must incorporate something like a symbolic formulation of what "computing" means. Thus no selection effected on the final output (if any!) would induce a drift, however slow, of the system toward the production of this mechanism if it were not already present in some form. Further, there is no chance ($<10^{-1000}$) to see this mechanism appear spontaneously and, if it did, even less for it to remain. Finally, we can predict what would happen if such a mechanism had been installed: for almost all the mutations the computation performed would have no relationship to the ones executed before: hence, no relationship to the selective pressure exercized on the output. All this, I repeat, is a simple consequence of the lack of matching between the space of the outputs and the space of the programs. This, of course, does not apply to the relationship between the space of param-

eters and adequate simplified models of the space of genotypes: They are theoretical constructs which have been specifically designed to fit. However, the question remains with respect to the relationships between the space of the chains of amino acids and the space of the organisms (or just as much, the parameter space studied by Sewall Wright). We do not know any general principle which would explain how to match blueprints viewed as typographic objects and the things they are supposed to control. The only example we have of such a situation (apart from the evolution of life itself) is the attempt to build self-adapting programs by workers in the field of artificial intelligence. Their experience is quite conclusive to most of the observers: without some built-in matching, nothing interesting can occur.

Thus, to conclude, we believe that there is a considerable gap in the neo-Darwinian theory of evolution, and we believe this gap to be of such a nature that it cannot be bridged within the current conception of biology.

Discussion

PAPER BY DR. SCHÜTZENBERGER

DR. ULAM: My impression is that what you have said so far is that one does not understand now how the blueprint determines the existing physical objects. That, of course, the Darwinians or neo-Darwinians would readily admit. Now, the assertion that such blueprints exist and are important is made much clearer through the discovery of the genetic chains as codes. Nobody in the 19th century or even now would profess to understand the details of how, from the code, an actual organism is produced.

DR. SCHÜTZENBERGER: We are not worried with the details. The only thing is that I would need an example where such a correspondence would exist or could exist, even in the simpler case.

The Chairman, DR. WADDINGTON: You have confronted us again, you have made

the gap because you have left out the middle space, the epigenetic space.

DR. WALD: What is epigenetics? What does the word mean?

The Chairman, DR. WADDINGTON: It is a derivative of an old Aristotelian word and means the study of the causal mechanisms of development. "Epigenesis" was used by Aristotle to mean that new things appear during development. Epigenetics is the name for the study of the causal interactions between the genes in the blueprint and the way they work together to produce first proteins, and then cells and membranes, myosin fibrils and God knows what. It is the causal study of the way the genotype space is translated into the phenotype and if you leave it out, of course there is a gap. Unfortunately, however, we can't yet put it

on computers! We really have got no analog
of development. This is why the whole ap-
plication of information theory to biology
breaks down; because what biological organ-
isms do is to treat information as axioms
and then develop theorems from them, and
this is something which isn't included in
information theory, as it is normally under-
stood. Information theory is a conservative
theory, in which information can't be in-
creased. But biology, as it were, starts with
Euclidian axioms and proceeds to write a
five-volume treatise on Euclidian geometry,
and it is that process which goes on in this
middle space of epigenetics and leads you
into the space of phenotypes. But nobody
has yet found how to do it on computers
and therefore, it tends to get left out.

DR. SCHÜTZENBERGER: I repeat, in order
to mediate between the space of chains of
amino acids and the real world of organ-
isms, some new construct has to be intro-
duced, and principles have to be stated ex-
plicitly explaining how this mediation is
conceivable.

At the level of molecular biology, we are
told that we have a reasonably complete
description of the mechanisms. Also, physi-
ology is providing us with an understanding
of organs. However, everybody seems to
take for granted that there is no gap in
between. I am not discussing the adequacy
of each of the two extremes. I just point out
that nobody seems to be able to give reasons
why they have anything to do with each
other. If there were explicit general prin-
ciples relating them, then we should be able
to simulate something analogous, and we
would have a lot of fun studying mathe-
matical models showing the passage from
disorder to order.

DR. ULAM: What you are saying, it seems
to me, is that the Darwinian and neo-Dar-
winian theories are not complete, and every-
body agrees with that; but it is not an ob-
jection to the scheme of things, which is
sort of lost sight of.

DR. RICHARD C. LEWONTIN: Can we give
you a practical experience where there is
no gap? Will that suffice? Suppose I tell you
that I know exactly the typographical
change involved in a mutation of the en-
zyme tryptophane synthetase. I know what

that change is and I know many such
changes cause an inactive enzyme to be
formed.

I know that an organism which is not fed
tryptophane, if it is an organism that re-
quires tryptophane in its proteins, will
not succeed in dividing and reproducing if
it has that typographic change. Therefore,
the frequency of such organisms will de-
crease in the population and be replaced by
those that can synthesize tryptophane.

Excuse me, but what step is missing in
this argument?

DR. SCHÜTZENBERGER: It is missing the
decisive step. Maybe I have too ambitious
a goal with respect to evolution theory; but
it seems to me that if its principles were
valid, we should then obtain on simplified
models the same type of correlation which
you claim to obtain. However, what we
know is that when we make changes of a
typographic nature, most of them are mean-
ingless from any respect, and when I say
"most of them," I mean less than one out
of 10^{100}.

DR. LEWONTIN: No, that is not true.

DR. ULAM: Tell him, Dr. Schutzenberger,
where his model fails.

DR. LERNER: Would you answer Dr.
Lewontin's question?

DR. SCHÜTZENBERGER: It is very intriguing,
but if you tell me that the coding is such
that this type of change induces meaningful
changes—what I mean by "meaningful" is
that they are related in one way or another
to external individual characteristics—you
already express a very strong hypothesis on
the living system. I say this is not included
in molecular biology as it is described now.

DR. LEWONTIN: If the speaker objects to
a case in which the enzyme has been de-
stroyed in its action, then I can give him
known cases where the enzyme, far from
being destroyed, is changed in its pH opti-
mum, changed in its isoelectric points,
changed in a number of aspects of its physio-
logical function by single substitutions of
single amino acids. We know exactly where
in the phenotypic topology of the protein
these amino acids have been substituted,
and we can specify exactly in what way they
change the physiology of the organism,
changing its fitness in the write-in space.

If you want, I can give you reference after reference.

DR. SCHUTZENBERGER: Yes, I can also give you references.

DR. LEWONTIN: As in hemoglobin.

DR. SCHUTZENBERGER: I can also give you a lot of anecdotes on typographic changes of books which transform some perfectly decent sentences into ones which are very funny to read in French.

DR. LEWONTIN: But, sir, I have said that they are not meaningless changes; they are changes that change the organism in its phenotypic optimum from one set of environments to another one.

DR. WEISSKOPF: I think the point Dr. Schutzenberger makes is the following: Dr. Lewontin is talking about changes in the enzyme by faults of reproduction; but Dr. Schutzenberger says that this is only a very, very small part of the typographic space and most of the changes seem to take place somewhere else in this space.

DR. SCHUTZENBERGER: I want to say that it is an observed fact that life works.

DR. WEISSKOPF: No, no, let's speak to the tryptophane!

DR. SCHUTZENBERGER: I want to know how I can build, on computers, programs in which —

The Chairman, DR. WADDINGTON: We are not interested in your computers!

DR. SCHUTZENBERGER: I am!

DR. BOSSERT: Perhaps I misunderstood, but I thought yesterday there was some discussion about a point quite in line with what you are saying. You have mentioned variation several times, and you require that a small variation at one level translate into a small, meaningful variation at the other. In fact, I think it came up several times yesterday that those instances where a small variation in the program space translates into a large variation in the phenotype space or output space, are usually of no interest.

DR. SHAHN: My understanding is that the problem involved is how you get from a lower organism to a higher organism; or at a different level, perhaps, what is it about the genetic material which is going to differentiate a horse from a pig; not how one

bacterial strain will die due to the deficiency of one enzyme.

However, the argument as it is presented, I think, could probably be leveled against all of biology in that, insofar as I know, there is not one mechanism which is completely understood. Any time there is a difficulty in getting from one step to another, an enzyme is introduced which often can be isolated, its properties in many ways expounded; but the mechanism is still left in a "black box". Using this terminology, one might be tempted to say that organisms have built in a "selectase," perhaps a "fitnessase," and these are part of an operon which is governed by "evolutionase." This now reduces all of evolution to the same state that most of molecular biology has been reduced to, and since molecular biology is today fashionable, I might claim to have solved all of evolution at the same level. I just have to isolate the enzymes.

DR. SCHUTZENBERGER: I want to make clear that my point is methodological, strictly methodological, and now we are just discussing facts. I am asking the question, How can you devise a program (or a book) such that typographic changes are meaningful? You are not interested in computers; I'm sorry. I am not very much interested in computers either, but here is an instance of a problem of order-disorder, and I am speaking of computers just to follow the Zeit-Geist.

How come that a system, which is not the type of system imbedded in the usual space-time topology, has the property that small changes within this typographic topology are meaningful? I could be specific here; I could document it with theorems.

DR. LEWONTIN: I gave you an example.

DR. SCHUTZENBERGER: I am not asking for examples. I believe you! But, I say, How come these changes are meaningful?

The Chairman, DR. WADDINGTON: He asked a methodological question. You have to answer it as a methodological question. He has asked, How do you arrange that typographical errors, changing letters and so on, have meaning when translated into this space? Surely, the way you do it is to have the typographic script set up as paragraphs with logical structure in the para-

graph. Most typographical errors will then be absorbed by the logic. You will see that there is a misprint and go on reading, understanding perfectly well what is happening. Occasionally, a misprint will change a key word into some other key word, and actually change the logical structure, but very rarely.

The main point is that your typographic space contains meaningful blocks. It is not a set of isolated individual words. It has meaningful blocks of connected connotations and this can absorb a great deal of typographical error but can occasionally have its meaning changed.

The very simplest case is where the epigenetic space is very reduced and you are just producing an enzyme; Dick Lewontin has pointed this out. In more complicated cases, like the mouse, where the development from the genes to a front leg is much more complicated, you have much longer blocks.

DR. SCHÜTZENBERGER: I am sorry, I want to challenge you somewhat. I am sorry to disagree flatly with the Chair. This is not an explanation but a postulate.

When you have said that there are meaningful paragraphs, you have already postulated the simplest thing, which is to make a computer program work at the paragraph level. For the time being there is no such possibility except by introducing beforehand the concept of meaning into it. There is dramatic change at the algorithmic level (that is the first time I have used the word but let it come now) between typographic errors of any sort and the ones which would preserve meaning. Taking paragraphs instead of letters is immaterial; it makes the case worse, that is all. So, what you say is all right except that what you propose is exactly the mechanism for which I am asking.

DR. BARRICELLI: The speakers seemed to stress very much the point that every step from one genetic pattern to another should be meaningful, but I think there is absolutely no requirement that every step should be meaningful. First,-you have many examples of changes indicating that often a large part of a protein molecule can be unimportant or play no role in its function.

You can lose a piece, you can add a piece, you can shift the reading frame in a segment of the RNA-molecule coding for the proteins, and so forth, and still leave the wild type function intact. That is one part of it.

Secondly, everybody can tell that by typographic change of the various types which have been mentioned today, you can change "Hamlet" of Shakespeare into Dante's "Divine Comedy," just by adding and subtracting pieces and changing one letter to another. So, if you don't require that every step in every place should be meaningful, you can make any large change you want.

DR. SCHÜTZENBERGER: I think there are two points here. First, it seems to me that this reckoning activity has some merit in that it shows that the matter is not that all changes must be meaningful. Only a reasonable proportion has to be. By "reasonable proportion," I would mean 1 10^{100}. It is not the case. We have a conflicting experience. You can quote me experiences where things work in life, but we have a conflicting experience in the computer. Although our processes are based on the same principles as the ones you state explicitly and the probability of a meaningful change is not one in 10^{100}, it is entirely negligible.

The second point has no relationship to the present discussion, but it has a more general bearing on these two days' discussion. It seems to me that it is a nice intellectual game to try to find that there is some path from A to B, but the problem is not to discover *if* there is at least one path. The problem is to decide if there is any reasonable chance of *finding* such a path; it is an entirely different question.

DR. LEWONTIN: I think I understand finally what Dr. Schutzenberger is getting at, and that is that the difficulty (and I agree entirely) is that most of the changes in a given environment would seem to alter the function in a way which is not, as you put it, meaningful. I think the thing that has been left out is the fact that we agree with this point, and it certainly is true, but that in a new environment the old messages, which had meaning in the old environment, now can no longer be called "correct" and changes can no longer be called "errors." On the contrary, as environment changes,

the present messages are no longer in a meaningful language and, therefore, new changes that occur are more likely to produce a real meaning in the new context.

That is the one point which I think all evolutionists are agreed upon, that it is virtually impossible to do a better job than an organism is doing in its given environment.

DR. SCHUTZENBERGER: I suppose we are getting nearer to an exchange of messages, but I was making a stronger point. When I said "meaningful," I was meaning meaningful in sort of an absolute sense. I say that all systems, similar systems that I know about, become meaningless in a radical fashion, when I make these sorts of changes.

DR. WEISSKOPF: In any environment?

DR. SCHUTZENBERGER: Yes, in any environment.

The Chairman, DR. WADDINGTON: This is not the case with the biological aspects.

DR. SCHUTZENBERGER: O.K. I have also the idea that something more must exist in biological systems, but the problem is to find the recipe so that we can simulate it on a different material.

DR. LEWONTIN: I think the answer is that you have over-estimated the number of absolutely meaningless changes that occur when you change a single nucleotide. If we list all single nucleotide changes and the known translation vocabulary between nucleotide triplets and insertion of amino acids, and then we list for a given protein all the results on that protein of changing amino acids all over the molecule, we will find, in fact, that a very large proportion of those do not render the molecule meaningless in an absolute context.

DR. SCHUTZENBERGER: You tell me it is factually all right. I ask you, What is the mechanism which makes it so, or what sort of conceptual mechanism could make it so? I don't know of any general principle or of any trick which in any other circumstances could produce this effect.

The Chairman, DR. WADDINGTON: Before we go any further, I think that, first of all, we should agree how we are using the word "meaningful." I think Schutzenberger means that when he changes something in program space, nothing comes out at all.

DR. SCHUTZENBERGER: It doesn't give support to any epigenetic effects.

The Chairman, DR. WADDINGTON: But actually when we change something, *some* protein does come out; it may not be a very good protein, but some protein comes out. All proteins do something, so all changes in the program level have meaning, in the sense that they produce a protein, except for some full stop marks, and so on.

DR. MAYR: Are you basically asking, why do molecules have such-and-such properties? Why are molecules the way they are? Is that really, basically, what you are asking?

DR. SCHUTZENBERGER: That's a good question. I don't think I have time to answer it now.

DR. LEVINS: I think the missing ingredient in this analysis is that you have left out evolution. The error of the reductionist methodology is to start out with a lower level and attempt to derive a higher level from it without considering the reciprocal relation. In fact, its topology is, itself, a product of evolution and we can start out with a given topology and describe how natural selection will modify this without knowing about the original underpinnings, especially as you get further and further away from the site of gene action to the interactions of these gene products. This is something which, in your library, would have to be described by 10^{10} simultaneous partial differential equations.

On an evolutionary level, in terms of some epigenetic parameters involving elasticity, the rigidness of the terrain and other things can tell us how the evolution is going to change it, the direction of homeostasis, of epistatic interactions, so that, in fact, the topology is the result of evolution.

DR. SCHUTZENBERGER: O.K., This is far easier to answer. You are falling into what I might call the Ashby trap. You only make the case worse by supposing that the mechanism which induces an agreement between the topologies has been produced also by random changes. That is to say, this sort of fallacy has been used a lot of times in "artificial intelligence" to pretend that one could write programs by machines which would learn how to tell themselves how to improve programs.

Still, then, at this level, the probability of meaningfulness is still slimmer by orders of magnitude which are 10^{100}.

If I had had more time, I could have dissected the typographic changes into three levels, each corresponding to a type of algorithm; each of them is practically irreducible to the previous one.

The Chairman, DR. WADDINGTON: Your argument is simply that life must have come about by special creation.

DR. SCHÜTZENBERGER: No!

VOICES: No!

DR. FRASER: Can I contrast one computer with another? You have a computer programmed to examine the statement, "All I am allowed to do is change letters and I hope I produce a program. Any kind of program will do." This doesn't work. We now turn around and set up another computer, and we tell it a basic genetic system of plus-minus alleles in which we are saying, "Can it produce information?" The decision on whether the information is useful will be a selective one of "survive or not survive." This is the same kind of decision-making; the programs look very similar to those which are being constructed to try to produce information-containing programs. The principles are very similar.

However, in the genetic one, the system is that there are multiplicities of pathways to suitable answers. The machine can gradually, step by step, get there; each step takes it toward the answers, and it produces them when all we have fed into the machine is a genetic system of essentially complete simplicity. What is surprising is how fast rational information is produced by the machine within the meaning of the original context.

So, if you are going to take a program space and say, "We cannot transform it," but leave out of it the means of combination and recombination in between and of evolution by selection, I am certain that your program will not produce sense; but if you put it in there the machine gets there so fast it is surprising.

DR. SCHÜTZENBERGER: What I have said is that insofar as principles have been explicitly stated, I have to deal with the whole space. To answer your question, one might believe that, in fact, life is using only extremely restricted subspaces of both spaces. What I am asking you, in all humility, is to provide me with a formal principle which would define those spaces, or to provide me with conceptual examples in which such spaces could be defined, even at the very modest level where they would have all the nice properties of matching. This has been done in a sense, at the Sewall Wright level: that is, on the space of parameters which by construction is correlated with the real world space. What I say is that such a type of restriction needs new conceptual tools, or principle, or what-have-you.

The Chairman, DR. WADDINGTON: I want Dr. Weisskopf to speak, but may I recommend that you have a talk in private with Alex Comfort; you can do it on his computer.

DR. WEISSKOPF: I want to analyze the difference of opinion between Schutzenberger and the rest of the world. This is, I think, the following: Schutzenberger says that in the typographical space, the overwhelming number of changes that can be done at random have absolutely no meaning, and he puts in support of it the fact that if you have a computer, and you change the program at random, it always is destroyed.

The other side says that that isn't so. The kind of program which genetics has produced with the 3-letter code is such that it isn't so. I think that is what Lewontin says, that a lot of changes, maybe not an overwhelming number but a large percentage, do make sense in the biochemical sense of the word, and here I think is the discrepancy.

DR. SCHÜTZENBERGER: There is no discrepancy. I am asking for you to tell me what principle to use.

The Chairman, DR. WADDINGTON: I regret we will have to leave this discussion at the moment; I think Dick Lewontin's is the next paper.