Information Theory in Physics and in Biology (*)

ERNEST H. HUTTEN

Royal Holloway College, University of London.

1. — Today we are in the midst of a new, conceptual revolution in science, the information revolution. The concepts of information and communication have become the main tools in the explanation of scientific theory: information flow rather than causal action represents now the basic model of a natural process.

There were three revolutions in physics since the beginning of this century, when the classical period had come to an end. Relativity theory (1905) was the first. Quantum mechanics (1927) was the second — for the original formulation of the quantum idea (1900) and the old quantum theory (1913) were not really a proper breakthrough; it was only with the Uncertainty Principle (and Dirac's relativistic version of the theory) that the conceptual revolution succeeded (1). Finally, the symmetry group theory of elementary particles is producing the third revolution today (2).

This modern period of physics has been described, for almost two generations now, also as «the crisis of causality». A new direction in physical theorizing was taken, away from causality as exemplified by the movement of mechanical particles towards a more «abstract» level of conceptualisation. It is indeed the concept of causality that is regarded even today by most people as the very foundation on which all science rests. First formulated by the presocratic Greeks, together with Atomism, causality acquired a precise, mathematical formulation with Newton's Laws. Ever since, the philosophical discussion of the concept has not ceased. I do not want to repeat here, once more, the endless and often futile arguments but look at the actual change in the use of the concept in physical theory. For this change that is still going on today is described by saying that the concept of causality is being displaced by the concept of information. We have reached the limit of the causal method in scientific explanation.

Newton's partial differential equation of the second order represents, physically interpreted, causal action: we have a path or orbit along which a masspoint moves under the influence of a force. The cause is represented

(*) Lecture held at the Istituto Superiore di Sanità on April 18th 1973.

Ann. Ist. Super. Sanità (1973) 9, 335-361
by the initial (and boundary) conditions of the movement and the effect is the position of the particle at a later time. For some generations, the nature of the action by the force was under discussion until action-by-contact was accepted, when the field concept and the finite speed of light had become established.

The metaphysics of causality played an important part in the development of physical theory and of science as a whole. In the end, it was physics and not philosophy that resolved the conceptual problems or, rather, abolished them when their «pseudo» character was recognised. We can briefly characterize these problems by the words «necessity», «certainty» and «absolutism». Physicists described these problems as resulting from the high degree of «idealization» of classical theory.

The «necessitarian» interpretation of causality found its formal expression through the mathematical rules of the differential calculus. It made explicit what countless, previous generations of thinkers had wanted, for psychological motives more than for reasons of scientific explanation. You may remember here the origin of the causality concept in the aitia of the Greeks: a moral conception that demonstrates the necessity of punishment for crime committed. The Ionian cosmologists achieved the first stage in «abstraction» by removing the human emotion from the idea and so they said that «nothing occurs capriciously, every process arises with necessity from a previous condition, or cause». Today, we would say that this statement enshrines the most important guide for scientific research, albeit in a metaphysical guise, namely the injunction «always explain». This rejection of miracles and supernatural forces was, historically, due to the Principle of Causality.

It was for this achievement — indispensable to rational thought — that the accompanying metaphysics took so long to dispel. Relativity theory was first with this «cleansing» operation. The absolute character of space and time was abolished and, together with it, the finite speed of light was recognised as the maximum velocity for any causal action. This means that causal action is limited: points on two different world-lines can be related to one another only within a certain space-time interval. There are causally indeterminate regions for any process. The causal chain is broken, consists of short pieces only, and their length depends on the reference frame chosen. (The Lorentz transformation easily demonstrates this behaviour of causal action). Absolutism is gone: causality depends on choice, to some extent. Necessity — or strict determinism — is curtailed. Moreover, reference to the chosen coordinate system, that is, the human observer, has to be openly made. The certainty of the result of any measurement can still be maintained, however; the classical fiction that all error is accidental and can be eliminated progressively without theoretical limit remains.
The next stage in the crisis of causality was reached with quantum mechanics (1). The Uncertainty Principle abolishes the classical ideal of certainty, that is, of the theoretical possibility to obtain error-free results. Interaction between phenomenon and instrument is, in principle, required to make a measurement. The interaction is limited by the uncertainties in the two conjugate variables — \( p \) and \( q \), or \( E \) and \( t \) — that are needed to describe the phenomenon. Thus, \( \Delta p \cdot \Delta q \sim \hbar \) and \( \Delta E \cdot \Delta t \sim \hbar \). Causality becomes statistical, therefore, since neither initial data nor final results can be specified except within the margin of error given by \( \hbar \). Causality is not abolished, as was first falsely claimed, because determinism is now impossible. Cause and effect, initial position at a given time and later space-time position are still related, though statistically. The main change is that interaction rather than one-sided action has to be accepted for causality.

Let us notice here that the Uncertainty Principle is formulated in terms of wave and particle since these are the carriers of causal action (2). The two sets of parameters are required to describe the movement of a particle or the spreading of a wave. But these two parameters are linked by the \( \psi \)-function which contains all the information we can possibly obtain since it is the complete solution of the Schrödinger wave equation describing the causal movement. Here the term «information» quite naturally, and even necessarily, occurs: neither «wave» nor «particle» are relevant when we speak about the result of a measurement. All the discussions about what kind of wave the \( \psi \)-function represents — «pilot wave», «matter wave», or «probability wave» — generated useless metaphysics because the mechanistic model of one-sided causal action was, openly or unwittingly, retained. Today we are all agreed that \( \int \psi \psi^* dq \) is the only meaningful expression; it tells us what happens by giving the probability of finding, say, an electron at a given place and time in an experiment. Interaction, uncertainty, statistics, and information are beginning to emerge as concepts needed on the more «abstract» level of microphysics.

Causality as a physical process represents the most general mechanism of how action is transmitted from point to point. Causality is thus energy flow. Quantized action, the stream of energy quanta, naturally can only be a statistical process, while mechanical action is visualized as a continuous push. How can we describe interaction, the mutual or reciprocal and, possibly, nonlinear relation between two physical objects?

The feedback is an instructive example to study for the understanding of interaction (4). The «causal loop» that leads back from the output to the input distorts linear and sequential, deterministic causality; it works through statistical causality. The control is actuated by the «misalignment» or deviation from a set standard; we have always fluctuations about the equi-

*Ann. Ist. Super. StAnlùa (1973) 9, 335-361*
librium state of the output. Instead of a direct, and single, link between cause and effect, we have a «transfer function» between input and output which is no longer determined by the input alone. The output is stabilized. Order and organisation is achieved, when information is sent back to the control that is set to a desired level of performance. Communication, or information flow, rather than mere causality describes the working of a well-regulated machine. Over and above the space-time sequence of events we have to consider the order and organisation of the system. The limits of causality are shown up when we have to deal with a situation that is being organised into a system as result of interaction between parts. Order as well as energy are then required to describe what happens.

When we arrive at the next, and at the moment last, stage in the development of physical theory — the symmetry group theory of elementary particles — to see any physical process as information flow rather than as causal transmission becomes more clear. First of all, the phenomena we deal with — the so-called elementary particles or resonances — are totally artificial, occur only in the laboratory and not spontaneously in nature. This is already partly true in the quantum domain, namely, in nuclear physics when artificial radioactivity and transuranic elements are produced. There is only one fundamental experiment, the scattering of a beam of particles off another such beam or a target, in which process the so-called elementary particles are created. Any phenomenon is called a particle that, in the experimental set-up, has no spatial extension but to which at least mass and charge can be ascribed. The elementary particles, moreover, have a very short lifetime — as small as $10^{-23}$ sec — so that their presence can be detected, sometimes, only from the behaviour of the more stable particles occurring in the same reaction. The basic information obtained from a reaction is in terms of the scattering cross-section whose value depends, naturally, on the coordinate system chosen, etc. We have, therefore, always interactions, not one-sided action of one active particle on another one at rest as is usually done in the classical domain. An experiment thus results in a census of particles and of possible reactions ($\dagger$).

First note that the whole aim of an experiment with elementary particles is classification, not the determination of a movement in space-time. Second, the classification is in terms of the symmetry of a group of transformations that are concerned with the quantum numbers of the particles. And, third, these symmetries are occasionally broken.

The «eightfold way» of Ne'eman and Gell-Mann thus leads, for example, to the multiplet of heavy particles, or baryons, of spin $\frac{1}{2}$ in the form of an octet. This exemplifies the symmetry of strong interactions under the group of transformations called $SU(\dagger)$. What are these basic symmetries?
Originally, in pre-relativistic physics, we have conservation of simple properties, like mass, momentum, etc. that must be obeyed in every reaction. The conservation guarantees that the property is «real». In relativistic physics, since any measurement depends on the reference system arbitrarily chosen, we replace conservation of properties by invariance of relationships under the Lorentz transformation which expresses the relativity requirement. Whatever remains invariant under a transformation to another Lorentz frame is «real». In symmetry physics, we have to consider many quantum numbers, for instance, spin, isotopic spin, baryon and lepton numbers, and strangeness; these relate to rotation, charge, right-left distinction, life-time, and mass. Thus, in a strong interaction, there is a group of Lorentz transformations, not only a single one, that determines the various quantum numbers. We have the famous CPT theorem stating that symmetry under the transformations of charge conjugation C, of parity P, and of time-reversal T should be preserved. However, certain reactions involving weak interactions, for example, violate the P-transformation, possibly even the C-P transformations; it means that the T-transformation must also fail since the CPT symmetry as a whole must stand (2). The details of symmetry physics are irrelevant here except for the logical, or semantic, scheme of a natural process that it entails.

There is a hierarchy of laws or principles; we have three levels. The symmetry principle leads to invariance which, in turn, leads to conservation; or group symmetry, invariance, and property are three successive levels of abstraction and description. Only property, within the classical domain, belongs to the causal scheme proper. Invariance, expressing a relationship, reflects the restricted causality of relativity theory. Symmetry expressing a classification has no longer a direct, causal meaning. Instead, symmetry expresses the order in which, by experimentation, we can place certain phenomena under certain conditions, e.g. the baryons. Classification on the meta-meta-level comprehends relationship on the meta-level and property on the object-level.

This has been also expressed by saying that symmetry is represented by principles, invariance by theories, and conservation by laws. Still another formulation is that of Wigner stating that the entities of one level become the objects of investigation on the next higher level. I think it is correct to say, then, that we have arrived in modern physics at a semantic hierarchy of conceptualizations.

Let me point out, too, that the age-old programme of physics — and of the whole of science through the impetus of physical concepts — which is atomism has come to an end with symmetry theory. The original Greek idea was immensely fruitful and, indeed, an inevitable and necessary basis for all science, namely, to go down to ever smaller pieces of matter and so

A Review of Scientific Instruments 1911 9, 335-361
to find its «ultimate» and imperishable constituents. It was the principle of explaining «the visible by the invisible» by interpreting the appearances in terms of an underlying mechanism. Until very recently, physicists believed that this programme would continue indefinitely; and the terminology of elementary particle theory illustrates this attitude. With atomism comes causality as, again, the Greeks had first formulated it. The picture of a natural process is then the causal encounter — movement and collision — between ever smaller particles. Elementary particles, however, except for the electron, proton and, possibly, neutron can not really be regarded as corpuscles: their short life-time and other properties like strangeness make this picture unrealistic. And the symmetry scheme suggests that we deal, on this level, with resonances either of a substratum — which is highly unlikely — or of a system temporarily formed by the collision of the more stable particles. The various strange particles are excited states — multiplets — of an artificially created, short-lived system in a similar way in which the excited states of an electron form a multiplet structure within the atom. We have come to the end of atomism though we are not yet quite clear about the possible programme by which we have to replace it on the so-called elementary particle level. One thing, however, is clear: we have been investigating, in physics, ever more complex and organised entities or systems, from simple masspoints to atoms, nuclei, and now elementary particles.

2. — This is illustrated, too, by the experiments performed. The evidence are some black lines — «stars» — on a photographic plate and, possibly, scintillations in a counter. These lines are interpreted as due to reactions between colliding and decaying particles. It is not their movement in space-time but the relationship to one another that is important here. Masses, momenta, charges, spins, etc. must be conserved. This conservation can, however, be established only by choosing suitable, relativistic coordinate systems to which the reaction is referred (the «laboratory frame», etc.). Finally, «good» quantum numbers are constructed so as to relate the participants of a reaction to one another by means of the symmetry of the group of transformations involved. Thus we arrive at allowed and forbidden reactions, e.g.: \( p + p = p + \Lambda + K^+ \) and \( p + p + n + K^+ \) (6).

The experiment itself, then, reflects the three levels of theorizing and although the evidence arises from the lowest, macroscopic, level of black lines on the photographic plate — which we may consider as causal — it is transformed by interpretation to explain another, and quite artificial, situation. Causality appears only as the blackening of the photographic grain through the action of the beam of particles from the synchrotron.
But we are not interested here in the photographic process and the silver specks formed. The relation between them — even the question whether they are particles or resonances — is not obtained from the direct, causal evidence. We are investigating a highly organised system, short-lived and strongly or weakly or electromagnetically interacting, and not isolated particles as suggested by the naïve interpretation of the tracks on the photographic plate, that is, on the «immediate» level of experience.

It is through the experimenter’s interpretation in terms of a theory that this is established. Or, the scientist imposes concepts to organise the evidence in a certain way and on a certain level, twice removed in fact from the actual photograph. First, he applies relativistic invariance, then he applies the symmetry of transformation groups; and so he arrives at an allowed reaction as formulated in an equation. This reaction shows how to arrange its partners — particles or resonances — into a symmetry scheme. This multiple interpretation of the evidence or the possibility of extracting several messages from the experiment is no more strange than, or different from the situation we find when we read a book: first, we have black marks on paper, then letters, then words, then sentences that convey meaning, etc. The artificial experiment in the laboratory involves active work on the part of the scientist. This is not merely practical but theoretical: modern experimentation requires technical skill but, equally, conceptual knowledge. We have to bring theory into the laboratory in order to project an experiment. Or, information about the organisation of a system rather than about causal movement becomes the purpose of research on the level of symmetry physics. I must content myself here to pointing out how close this attitude is to that in a «living» science when, for instance, we investigate an organism (7).

It is normal talk among physicists to say that classical physics, relativity theory, quantum mechanics and, now, symmetry group theory are constructed from conceptions in an ever increasing order of «abstraction». We take the level of ordinary, macroscopic experience as standard since we directly work on it in the laboratory: reading a scale, touch, smell, sound represent the «concrete» basis to which, in the end, all our theories must relate. Newtonian mechanics, although close to ordinary experience, is already an abstraction since it is based on the concept of a point particle. Successive theories — of relativity, quantum and symmetry — made the concepts of particle and wave, space and time more and more abstract (and introduced new concepts as well).

In the same way, however, the role of the observer or experimenter has to be more and more acknowledged. The relativistic observer and quantum-mechanical experimenter have to be included into the respective theories; and this is the more true of the symmetry group theory. In Newtonian
mechanics, the «human element» is totally excluded from the theory — the idealization is greatest on the most «concrete» level. Certainty, absolutism, and the necessity of deterministic causality are possible conceptualizations on this level. Idealization is decreased and «abstraction» and the experimenter’s participation in the measuring process increase when we advance to the newer theories. With it, the conception of information transmission as the model of a natural process arises; while causality decreases with abstraction — it is part of the «concrete» model. We know from the history of ideas as well as from psychology that the concept of causality first arises in infantile experience and is then gradually refined for scientific use.

There is an interaction between experimenter and Nature, between instrument and phenomenon by which information is created. This more realistic description of the process of measurement brings with it the Uncertainty Principle. If we look at any of the «thought experiments» that illustrate quantum-mechanical measurement — the γ-ray microscope or the double-slit — in what does the measurement consist? We shoot a photon at an electron which is localised at q in order to determine its momentum p; but the photon energy knocks the electron out of position so that \( \Delta p \cdot \Delta q \sim h \), etc. The measurement consists in registering an energy and in evaluating a position; the same applies to the other, possible process, e.g. evaluating a time period so that \( \Delta E \cdot \Delta t \sim h \). Any measurement is an act of comparison; a scale is imposed and we have to read a deflexion. There are two activities involved then, only one of which — registration — is concerned with energy, hence causality. For causality is the transmission of energy, an action-by-contact. The second activity of the experimenter — evaluation — is made possible only by the previous preparation of the experiment. We have put a metre-stick — or, in the γ-ray experiment, a microscopic scale of some sort — in the right place within the experimental equipment so that we can read it off. Such a scale-reading — the deflexion of a needle relative to a scale — is the prime constituent of any physical measurement. The movement of the needle indicates that the event desired has occurred, that is, the causal action, while the exact amount of the needle’s deflexion allows us to evaluate the strength of this action. The scale — always relative and even to some extent arbitrary — allows us to order the causal action in terms of energy or some other related parameter. Both energy and order, therefore, are needed if we want to obtain complete information from an experiment.

The order, however, is introduced from the outside, so to speak: it is not part of the transmitted causal action. Within the causal picture, all we have is an energy link between two space-time points or events. There is a causal order — simply the series of earlier to later events. This
is not the order we require for the knowledge of a measurement; we need then to know how to order the strength of the causal action, that is, of the effect we are investigating. Order is always a second-level property. The order of the causal sequence is found when we speak *about* the space-time positions and making them into a series of numbers. The order involved in evaluating a measurement consists in speaking *about* energies (and positions) and arranging them into a series within a certain margin of error. In both instances, we are on a semantic level one higher than that of the objects — either space-time positions or energy-values — which we put into an order.

Thus, measurement is an *information process*, the centre of the whole activity of experimentation. The classical tradition in physics makes us exclude the experimenter from the result of his activity in order to be «objective». Thus, an important and indeed essential, aspect of what he does remains outside the theory and appears only as part of his practice; it is unformalised and merely laboratory procedure. This account of scientific method is adhered to even in quantum mechanics although, there, we take interaction rather than one-sided action as basis for measurement. It is still described as a causal process concerned only with energy transmission in space-time, though statistical rather than determinist. We cannot speak of orbits within the atom any longer — which really is a serious break-down of the causal picture — but the description of an atomic process in terms of energy levels saves the causal method. What is not recognised is that the Uncertainty Principle is indeed a principle — one semantic level higher than, say, a conservation law — since we speak *about* $p$ and $q$ and evaluate their errors.

The information we obtain from a quantum-mechanical measurement is not openly specified. The experimenter, in spite of the interaction of phenomenon with the equipment, remains a passive outsider content with mere registration although, at least, the «disturbance» due to it is now theoretically acknowledged. We have still only energy transmission and (statistical) causality. Over and above the quantum-mechanics proper describing phenomena we have, however, a *theory of measurement* based on the Uncertainty Principle and describing the statistical ordering of results (?). This is really a theory concerned with an ideal experimenter's activity which is expressed in terms of a measurement operator. When such a theory is looked at in detail, we discover a striking resemblance with information theory. The $M$-operator introduces an order in the results obtained according to the evaluating criteria of the experimenter. If the amount of information is represented as a vector in function-space, its transmission — input to output of the communication system — is described by a transfer matrix $\langle x| A | y \rangle$. From this basic formulation,
assuming suitable constraining conditions such as unitarity, mathematical expressions are derived that encompass both signal transmission and quantum-mechanical measurement.

The emergence of the concept of information replacing causality as the « mechanism of Nature » is then parallel to the acknowledgement of the role of the experimenter and to the increasing abstraction of the theory. Or, instead of naively seeing Nature as an independent and fully constituted entity that we can only investigate from the outside, we now see scientific research as an activity directed towards the purpose of gaining information. There is an interaction between experimenter and Nature by which information is created; but we have to design an experiment, that is, start with some knowledge before we can extract new information. We deliberately « put Nature to the test », as Bacon said; and, after all, active experimentation heralded the beginning of modern science. The more we acknowledge science as a human activity, the more information rather than causality becomes its key concept.

3. — Information theory was first formulated within the context of electrical communication engineering (9). The similarity of Shannon's formula to the expression for entropy in classical thermodynamics was seen at once. Thus, the context of discovery set the telegraph model for the transmission of information. In this model, both sender and receiver know that the impulse, or energy, sent is a message, that is, that it conveys linguistic meaning. They usually even know the code in which it is transmitted for the purpose of easier communication. The flag code of naval signals or the Morse code of telegraphy are the example. Thus, the only problem is to receive the impulses of electric energy as correctly as possible, as free from accidental disturbance, or noise, and distortion as can be achieved. There is no doubt as to whether or not meaning is conveyed by the message: it is only the question of what meaning, because the signal might be distorted. What is transmitted in the telegraph model is an array of symbols whose meaning is known beforehand and information is then only a measure of their complexity, not of their content. Both sender and receiver know the same language, employ the same code; meaning is presupposed and outside the theoretical description of the information transmission.

This is the « classical » information concept. Its limitation has been widely recognised and information theory has therefore not found the appreciation it deserves. Gradually, however, a widening of the concept has come about through its application in molecular biology and in the pattern recognition of automata (10).

In the DNA model of genetic information transmission we do not know whether or not the process represents a communication at all.
process consists of at least three stages: DNA duplication, transcription into m-RNA and t-RNA, and translation into the array of amino acids which results in the protein synthesis. We have to interpret the sequence of events, establish an order, and recognise the meaning of that order.

Meaning enters here openly: we can have sense, missense, and nonsense in the message, that is, the translation may result in the correct protein, or the wrong one, or none at all being synthesized. The meaning is, as always, by «interpretation», a process that is most easily understood in linguistic terms. We have to speak about a phenomenon or sign — the black cloud, for instance — when we want to interpret it as threatening rain. Smoke means fire; the natural sign becomes an artificial symbol by human interpretation in a given context. A symbol has a meaning that is added by us to the physical sign for a purpose; and we have rules for doing so. «Cavallo» means horse is an example illustrating that meaning is given by rules of usage. These rules specify the range of possible cases in which the word applies; and these possibilities are based on our experience of phenomenon or situation involved. There is only a selective, limited range; and it is the choice from the finite set of possible uses that gives meaning to the word in a particular instance.

Information flow conveying a message is both energy and order. We need to transmit, in the general case, the meaning as well as the amount of information. The genetic model of DNA transmission allows this, which the telegraph model cannot do. A linear array of molecules — the base sequence of DNA — becomes a series of symbols when it is transcribed and translated into the amino acid arrangement of proteins. Replication — or energy transmission; transcription — or coding; and translation — or the specification of meaning, are the three steps involved. Chemical valency and enzyme action effect the transmission.

The DNA alphabet, as we know now, has only four letters and the information stored in the molecular chain is contained in a three letter word, the codon. There are then $2^3 = 64$ three letter bits of information that relate to twenty amino acids. We have a «dictionary» from which we find the «word» for a particular amino acid. Each sequence of words is a meaningful sentence, being the gene that specifies a particular protein. In this way we interpret the DNA chain as a meaning carrying arrangement: we make the synthesis of proteins the meaning of the sequence of the two nucleotide base-pairs. A triplet like UUU — uracil or thymine, according to whether we use DNA or RNA — corresponds to phenylalanine, etc. The analogy with language can be pushed even further if we consider the transmission of the genetic information in more detail, for example, investigate the role of the structural genes, operons, etc. We can then identify sentences and paragraphs in the DNA language. The exact details of the
process are not, at the moment, important for my argument. What we need to see is that the physical order, or organization, of a number of ATCG molecules in a fairly long stretch of the chromosome represents the message carrying an instruction, or potential information. The meaning is given by the order — first, of the molecules themselves forming a triplet and, then, of the sequence of triplets — for, according to the order, a different amino acid and, eventually, a different protein is produced (11).

Thus, in contrast to the telegraph model, the DNA model does involve meaning. Two levels of activity are involved. We have a pattern that conveys the information much like the ideograms of the Chinese language do or used to do when the language was in its more primitive form. In the telegraph model, the existence of the pattern is known a priori, while in the genetic model it has first to be found. Thus, the second-level activity — translation — has to be openly referred to and even the coding process — the transcription into m-RNA and t-RNA — has to be scrutinized, while the Morse or flag codes are fixed and known before the transmission starts.

There are then two levels of entities involved — the actual chemicals A, T, C, G; and their triplet grouping in a sequence, or the order and organisation (it is organisation rather than order since we deal with a system). The relation between the two levels is one of interpretation; it is to some extent similar to radio transmission in which the carrier wave — the energy — is modulated by the imposition of the human voice, that is, of order.

At the same time as the discovery of the double helix, von Neumann had demonstrated the need for two levels of entities in the construction of self-reproducing automata (12). This is the same problem, the transmission of information from one generation to another. The false antithesis of epigenesis vs. preformation, which arises from the causal model of energy transmission, illustrates the need for the information model as well.

Von Neumann’s account is now well known. There must be new material which a self-reproducing machine can put together by template action if the machine is surrounded by a «soup» of suitable parts. And there must also be a description of the machine which the automaton can read without interference to its operations (this is the «passive» method that avoids the logical paradox of the Richard type occurring when the automaton examines itself and freely constructs a description of itself, in the «active» method). This description on a tape, like the double helix, can reproduce itself and so be added to the offspring, etc. That is, apart from a template, there is also a blue-print. The constructing automaton will interpret this instruction by assembling the parts in the supply to make a new automaton.
The theory of von Neumann illustrates the difficulty by which any discussion about meaning is bedevilled. We are so used to think of the meaning of words that we forget the possibility of non-verbal meaning. Whenever we have one thing that stands for another thing in a certain context, whether a "natural" sign or an "artificial" symbol, we have meaning; we interpret the sign or symbol guided by a theory and by past experience, that is, by rules. The blue-print, or the tape describing the automaton, are on the verbal level. The blue-print for the protein synthesis is the order of the base-pairs; it is "immaterial", so to speak, but carried by the nucleotides as the modulation is by the carrier wave in radio transmission. This coded instruction is one level higher than the molecular one, a meta-level relative to the object-level of the base-pairs. Meaning is not just simply order but the recognition of order. The ribosome moves along the tape of m-RNA and "reads" it when the anti-codon of t-RNA "recognizes" the codon of the m-RNA. While the template produces order, the blue-print specifies and recognizes it. This is non-linguistic communication, more primitive but also more basic than ordinary language. This communication, over and above causal relationship, is the process characteristic for evolution (13).

4. — The codon-anti-codon mechanism shows that the reception of information is more like pattern recognition; and in a pattern, the meaning is shown by the shape. We must distinguish here between the source — the stored information, the transmission process — which requires transcription, initiation, termination, etc. — and reception or recognition of the message. Again, in the telegraph model, we tend to obscure the difference between stored and received information since, apart from noise, we assume the same probability distribution for the symbols in both of them. In spite of possible distortion, there is a link between the two kinds of symbols, for instance, we speak about the mutual information $P_{ij} = P (b_j \mid a_i)$ when symbols $a_i$ are sent and $b_j$ received.

Pattern recognition, however, has two independent, semantic dimensions. First, the signal must be received and registered; and, then, it must be classified and judged whether or not it fits into a prescribed range; in this way, the meaning of the message is established. In the most general case, reception is effected through a trial and error method. This leads at once to a generalisation of the usual information (or entropy) formula.

The usual formula for the amount of information is $H = - N \sum_i p_i \log p_i$, which expresses the average of a diversity of $N$ similar messages with probabilities $p_i$. This is the variability, as Bonsack called it (14). If we do not average, however, we can find the specificity of the individual messages. While variability refers to the potential set of messages, the specificity
describes the individual, actual messages. Thus, in the general case in which the probabilities of each symbol of the message are different, we have \( H = - N \sum f_i \log p_i \), where \( f_i \) is the relative frequency of a given symbol \( X_i \) in a particular complexion. This is a straightforward derivation from Boltzmann's principle. We have two different quantities \( f_i \) and \( p_i \) to consider, which become identical only «in the long run» when the relative frequency approaches the probability of occurrence of a given symbol.

How can we exploit this generalisation in order to describe the reception of information and pattern recognition? This more general form of the entropy formula appears obvious and natural to any physicist who has ever discussed the standard form. There is no reason why we should always average and assume that the probabilities of the symbols sent and received are the same. When the transmitter and receiver employ different «alphabets», or «categories», etc. — not previously agreed upon by convention as in the telegraph model — then two different probabilities must occur. While the sender emits symbols with the probability \( p_i \), the receiver assumes a probability \( q_i \) for any symbol \( i \) transmitted. In the absence of any previous knowledge, the receiver keeps on changing his probability distribution and if he manages to make \( q_i = p_i \), then maximum information will be obtained. This is a natural generalisation of the communication process.

The receiver then has a set of categories or classifications \( A_i \) such that the uncertainty for each is \( N(A_i) = - \log q_i \). Now, the symbols sent are distributed with \( p_i \), and so for all \( i \) the total uncertainty is \( N(A) = \sum_i p_i N(A_i) = - \sum_i p_i \log q_i \).

If the receiver were only registering a signal, then only one kind of probability, \( p_i \), is needed, as in the telegraph model. When the receiver has to evaluate the message, or recognise the unknown symbols, or find their meaning, then he has to classify the symbols in some way, and this introduces the second probability \( q_i \). Any two-level process must require two sets of probability, and so we can write the most general formula as \( H(p/q) = - \sum_i p_i \log q_i \). This is the uncertainty of a message when the symbols are sent with probability \( p_i \) but the receiver employs the probability \( q_i \) for deciphering it. It is then clear that the uncertainty \( N(p/q) = N(A) \) reduces to the usual formula for \( q_i = p_i \). Thus \( H(p) = N(p/p) \).

Recently, Bongard (6) has also developed the same formula when describing the pattern recognition process as problem-solving activity. His arguments are very general and powerful and he showed, too, that we have \( N(p/q) \geq N(p/p) = H(p) \). He has been able to construct a comprehensive theory of pattern recognition which, to my mind, very largely answers all important questions. One particularly interesting concept Bongard introduces is that of useful information.

*Ann. Ist. Super. Sanità (1973) 9, 335-361*
Assume a finite set of patterns $M = \{m_k\}$ that occur with given probability distribution $p(m_1) \ldots , p(m_k) \ldots p(m_t)$. This set is subdivided into $n$ subsets $A_i$ that are mutually exclusive and together exhaustive of the problem: $A_1 \lor A_2 \lor \ldots \lor A_i \ldots \lor A_n = M$ and $A_i \land A_j = \emptyset$, when $i \neq j$.

The solution of the problem $A$ is to find some patterns such that $m_k \in A_i$. Assume a solution algorithm that chooses a pattern $m_k$ and a subset $A_i$, perhaps at random. Assume a testing algorithm that can determine whether or not $m_k \in A_i$. If the answer is yes, the problem is solved; if not, the selection of subsets proceeds, etc. The solution algorithm determines the probability distribution $q_i$ for the selection of the $A_i$.

The number of times the testing algorithm has to be applied before a true answer is obtained is the problem uncertainty that we take as the log of the mathematical expectation of the number of trials, which is a function of $p(m_k)$. Consider now a system consisting of solution algorithm $S$, communication channel $C$, and a decoding algorithm $D$ connecting the two. For every value of the incoming signal $C_k$, the D-algorithm substitutes a new value $q'_i$ for each $q_i$, which produces a new probability distribution for the solution algorithm. For example, the decoding algorithm removes the old roulette wheel and puts in another one with different sectors; therefore the incoming information changes the problem uncertainty. Suppose the problem had the uncertainty $N_0$ for a given solution algorithm prior to receiving information from the communication channel, and $N_1$ afterwards. Then, the amount of useful information transmitted is $I_a = N_0 - N_1$. This makes the problem of pattern recognition thoroughly relative as, indeed, it must be. For we can speak about useful information only if we know the problem, the initial state of the solution algorithm, and the properties of the decoding algorithm.

The change in problem uncertainty due to the incoming signal can be seen as the process of accumulating knowledge in the form of the probability distribution $q$. We assume the useful information to be zero when $q_i = 1/n$ (the optimal distribution if we know nothing about the $p_i$). The increment in useful information contained in hypothesis $q$, with respect to the answer probabilities $p$, is given by $I_a = \log n - N(p/q)$.

A great deal more can be said about the concept of useful information and pattern recognition. Bongard (*) has worked out, too, a method of finding the problem uncertainty if the signal is transmitted along a channel characterized by $\|P_{ik}\|$ and decoded by an algorithm given by $\|d_{ik}\|$. A formula is derived establishing a relation between the properties of the problem, of the signal, of what the observer thinks before the signal is received, and of the benefit he derives from it under the circumstances. What we do need to say is only that the method of pattern recognition here outlined

\[\text{Ann. Ist. Super. Sanita (1973) 9, 335-264}\]
solves the problem of meaning, at least in so far as transmitted or received information is concerned. Another approach is proposed by Saracevic (2) who defines «semantic information» in terms of a thesaurus $\Theta$ in which our previous knowledge is recorded. This is very similar to the problem set $A_t$. The amount of information $I(\Theta, T)$ is then defined as the degree of change of $\Theta$ when statement $T$ is received. Again, we have an algorithm and rules according to which decisions are made and so meaning is established.

There is one remark I would like to make. The solution and testing algorithms are really formal representations of the evolutionary mechanism of random variation and natural selection. Items are selected at random; and then they are compared with a standard or fitted into a category or class. The latitude, or margin of error, within which an item is accepted shows up the basic uncertainty that characterizes the information process.

5. — Meaning is inseparable from order and organisation; and organisation implies the existence of a system. Interactions, including non-linear ones and feed-back, between the parts of the system are responsible for this organisation; and different, organised levels may arise according to the interactions involved. Moreover, there are boundary conditions that specify the behaviour of the system as a whole. Finally, interchange of energy and entropy or information with the environment cannot be excluded, certainly not for a living system. There is metabolism, and the organisation of the system may change in time. The range of possible behaviour of a system is fixed by the constraints imposed on it. Take as example the vibrating string: its possible modes of vibration are determined by the manner in which the two ends are fastened, i.e. the boundary conditions; and different tones can also be produced by the manner in which vibrations are excited from the outside, i.e. the initial conditions. Thus, the idea of system — of an entity made up of many, interacting parts that is separated from the environment, however temporarily and even partially — is indispensable to meaning.

A physical system, for instance, of thermodynamics is characterized by its state variables; it has energy states and the transitions between them represent the causal process that the system is carrying out. Order, in the form of the state variable of entropy, enters into the description of the process; but entropy is accounted for in terms of energy, for instance, when its value is multiplied by the temperature which is the main concern of thermodynamical theory. A communication system, however, must be more complex simply because flow of information rather than of energy alone describes the process. A living system offers therefore a better illustration than a strictly physical system that may be orderly but lacks

Anna. Ist. Super. Sanita (1973) 9, 335-361
complex organisation. The hierarchy of levels of an informational system must be specified by the (actual and potential) amount as well as by the meaning of information.

In his theory of self-reproducing Automata, von Neumann \(^{(12)}\) took complexity as the characteristic, a concept he himself regarded as vague and unsatisfactory and, ultimately, to be explicated in terms of information. There are two possibilities, opposite to each other in tendency, when we look at self-reproduction of living organisms or of automata. Either the parent is more complicated than the offspring — as you would see it in the machine producing another machine — or the other way around — which is exemplified by the evolution of life that is in the direction to greater complexity. Today’s organisms are phylogenetically derived from others which were vastly simpler; it is inconceivable how any kind of description of, or instruction for, the later, more complex organism could have been contained in the earlier one. For the self-reproducing machine, the argument leads to the opposite conclusion. A machine tool is more complicated than the parts that can be made with its help. Generally, the automaton \(A\) that can make another one, \(B\), must contain a complete description of \(B\) and also rules as to behave while effecting the synthesis. Here, complication, or productive potentiality of the organised system as von Neumann called it, is degenerative. He drew the conclusion, however, that complication is degenerative only below a minimum level of organisation. The artificial production of a more complicated from a less complicated object is, in principle, possible. But it requires a flexible, self-correcting programme — stored on tape, for instance — that the constructing system must interpret in accord with external conditions. The system must be sufficiently complex to have the ability, as I would say, for recognition and meaning.

If you look at von Neumann’s principle of complexity from the viewpoint of thermodynamics alone, then what it seems to say is this. We know that the free energy \(F = U - TS\) determines everything a thermodynamic system can do; and the actual process goes so as to minimize free energy. Only at absolute zero, we have \(F = 0\), or \(U = TS\), the internal becomes equal to the bound energy; and a process can then go on without loss. The bound energy represents disordered energy, energy that has lost its directional quality, or the ability to do work, as we say in physics. If that bound energy outweighs the internal energy, no process can occur.

If we see now information as energy plus order, then a similar tug-of-war could be envisaged for communication that exists for the thermodynamical process. However, what corresponds to «internal energy» is here «stored information», or «redundancy», while the noise in the system represents the «bound energy». Uncertainty, or potential information,
plays the role of «free energy» that must have a positive value for information flow to occur. (It is easy to be misled by the terminology we employ and identify, wrongly, bound energy with stored information, as I have done some time ago)\(^{(16)}\). A system, whether an automaton or a living organism, has to be constructed or grown from previous knowledge, whether this knowledge is of physical or of genetic law. If it is to be self-reproducing, the system must be sufficiently organised to have a self-correcting programme, that is, the programme must contain sub-programmes which can be carried out when the need arises. The programme must therefore consist of at least two semantic levels of organisation; meta-levels, meta-meta-levels, etc. are needed. Such complexity of organisation can be imagined to arise from interactions, especially through feedback, since this kind of interaction is actuated by fluctuations or noise. Moreover, feedback or retroaction tends to make an effect independent to the degree to which it acts on itself; a stable state is reached, and on this level of organisation we find a kind of self-determination. An integration of the system is achieved through the interaction of the system’s internal with some external parameters, as P. de Latil has so cogently demonstrated a long time ago\(^{(1)}\).

Thus, we need a large amount of stored information, or a high degree of redundancy but, equally, a lot of uncertainty or noise in the system if it is to have sufficient degrees of freedom for self-reproduction or novel information. There exists again a conflict or polarity, this time, between stored information and noise. If we reduce the noise to nothing — if that were possible — we could not generate new information and the system would be dead. A thermodynamic system at absolute zero, however, exhibits the maximum degree of order and functions most efficiently in the absence of entropy. We must, I think, distinguish strictly between organisation and order: a crystal exhibits a high degree of order but contains little redundancy, while in a communication system organisation and redundancy increase (or decrease) together; and such a system must contain noise. One might say, in a somewhat paradoxical way, that polarity, uncertainty, and information characterize the organised system while dichotomy, certainty, and causality apply to the idealized, physical system.

Another, speculative remark: according to von Neumann’s principle, we have no self-reproduction (or information flow) if the complexity of the system is too low. It seems natural to say that there must also exist an upper limit to the complexity, or degree of organisation, of a system. An organised system must have both redundancy and noise. Since both increase together, considering their «polarity», that is, the uncertainty principle of a sort that relates them to one another, we could imagine that the

\[\text{Ann. Inst. Super. Scienp. (1978)} \ 9, \ 355-361\]
relative difference between them becomes too small and the system un-
viable. The system might «explode» if the stored information can no
longer contain the noise.

Redundancy is expressed by the Shannon formula \( R = 1 - \frac{H}{H_m} \),
where \( H \) is the actual and \( H_m \) the maximum amount of information. It
is zero when actual and maximum (potential) information agree in amount,
for the ideal channel capacity. The channel plus correct coding set the
limit to possible information. The ideal, noiseless channel is the zero-
point. Redundancy is usually described as the capacity of the information
source to overcome error. There are, however, at least two kinds of redu-
dancy, by repetition and by rules. The simplest but, obviously, most inef-
fective way is to repeat the message over and over again in order to
insure correct reception.

The rules that may link the symbols of the message provide a more
efficient method though they must be known to, or at least discoverable
by, the receiver. The rules of language, or grammar, are the example;

in the logical point of view, we divide them into the formal rules of
syntax and the rules of interpretation, or semantics. The point I want
to make here is that redundancy introduces organisation, or structure,
into the information source. This organisation gives the basis for meaning
of any possible message though it will depend on our knowledge of the rules
whether or not we recognise it.

It is important to distinguish clearly between stored information of
a system and the transmitted information we can get out of it.

Again, the telegraph model tends to obscure the difference since it is
mainly concerned with the transmission process and both the structure
of the source and the recognition of the message are judged from this point
of view. Redundancy is therefore related to combatting the noise of trans-
mission rather than to establishing the possible meaning that the information
source can provide.

Redundancy is given by Shannon’s formula as the difference of the
relative entropy from unity which stands for 100% or total loss of infor-
mation. The same formula can also be written as a difference, \( R \cdot H_m =
= H_m - H \), which brings out more clearly that redundancy is both relative
and arbitrary, in the sense that the zero-point of the scale has to be fixed.
We have an entropy scale, and entropy is physically measurable only apart
from a constant, that is, as a difference; we usually set its value as zero
at the absolute zero of temperature as the Third Law of thermodynamics
allows us to do. If we look then at the sequence of symbols, two natural
reference points for the redundancy can be specified which both refer to
the state of maximum entropy: equiprobability and independence of the
symbols in the message sequence (17).
For a sequence of single symbols, we have the entropy 

$$H_1 = - \sum_i p_i \log p_i$$

(using log to the base two and binary units). The entropy has a maximum when all \(n\) symbols are equiprobable. Then \(p_i = 1/n\) and \(H_1 = \log n\). The difference from the equiprobable state is given by \(D_1 = H_1^{\text{max}} - H_1 = \log n - H_1\).

For a sequence of connected symbols, let us consider first only the conditional probabilities that link two successive symbols. This is the simplest case. In the usual \(P\)-notation, this is expressed by the formula \(p(AB) = p(A) \cdot p(B|A)\). In terms of a probability distribution, we have \(H_{2}^{\text{le}} = - \sum_{i,j} p_{ij} \log p_{ij} \) and therefore \(H_{2}^{\text{le}} = - \sum_{i,j} p_{ij} \log p \cdot p_{ij} \) if the symbols are independent. Obviously, \(H_{2}^{\text{ne}} = 2 H_1\). And \(H_{2}^{\text{le}} = H_1 + H_M\), where \(H_M = - \sum_{i,j} p_{ij} \log p_{ij} \). The divergence from independence is then defined as the difference \(D_2 = H_2^{\text{ne}} - H_2^{\text{le}} = H_1 - H_M\). The sum \(D_1 + D_2\) measures how much the entropy has been lowered from its maximum as result of redundancies; this represents the stored information, so that \(I_i = D_1 + D_2\). We also have \(H_M \leq H_1\), a formula first given by Khinchin (1957) showing that order increases with dependence (17).

The theory of redundancy is greatly extended when we recognise that a series of connected symbols represents a Markov process. This is well known from the telegraph model in which an information source is analysed in terms of the Markov transition probabilities between states. This model, however, is concerned with the transmission process mainly and so the import of Markov probabilities for stored information, as a characteristic of redundancy, was not seen. It was recognised only recently by Gatlin whose account I am following here (19). We say then that independence represents a zero-memory sequence; conditional probabilities linking two successive symbols are described as a first-order Markov sequence, \(m = 1\); for \(m = 2\) the probability of a certain symbol occurring in the sequence depends on the two preceding symbols, etc. The \(m\) preceding symbols are the «memory» due to stored information. The general \(m\)-order Markov process allows us to describe a system, or sequence of symbols, of great complexity and allows us to introduce a hierarchy of levels of organisation.

In analogy with \(D_1\) and \(D_2\) we have \(D_{2}^{(m)} = H_1 - M_{M}^{(m)}\) and there are also higher-order divergencies \(D_l = H_{l}^{(M - 1)} - H_{l}^{(M - 2)}\), where \(l\) is always one unit higher than the maximum memory index. Thus we can define \(D_{2}^{(1)} = H_{M}^{(0)} - H_{M}^{(1)}\), where \(H_{M}^{(0)}\) is a more consistent notation for \(H_1\). This first increment of \(D_{2}^{(m)}\) is followed by \(D_3 = H_{M}^{(1)} - H_{M}^{(2)}\), and \(D_{2}^{(2)} = D_1^{(1)} + D_2\). In general, Gatlin finds \(D_{2}^{(m)} = D_2^{(m)} + D_3 + D_4 + \cdots + D_{m+1}\). If \(m = 0\) is included, \(D_{2}^{(0)}\) denotes \(D_1\) and the total diver-

---

*Ann. Int. Super. Sanitè (1973) 9, 335-361*
The theory of redundancy is the key for our understanding of organised systems and organisms if we apply it to stored rather than to transmitted information. The concept of redundancy requires more analysis than it has been given hitherto. It represents not only the capacity to overcome noise: redundancy as a measure of stored information, or previous knowledge, also can serve to indicate value, or meaning, or complexity (to use von Neumann's term). Some time ago, I tried to express the value of a message, in contrast to its amount, by what I then called «complementary information». This, too, was a difference from the random state given by 100% redundancy, similar to $D$ mentioned here. However, I kept to Shannon's relative entropy and this prevented me from seeing clearly that equiprobability and independence have to be separated; I remained too much within the ambit of «classical» theory. The theory of Gatlin is both elegant and a very great advance in our understanding of redundancy and meaning.

6. — I want to discuss briefly what the DNA model allows us to say about the hierarchy of levels of an organised system (7). The Markov process represents memory — the history of the sequence preceding the $m$-symbol block. The interdependence of symbols is introduced by constraints which are the redundancies prescribed by stored information. The constraints are imposed on each level from outside; they determine possible behaviour of the system on this level. Since meaning is mediated by relationship and given by rules, constraints or stored information allow us to classify symbols and assign meaning to them. Meaning is always established only by reference to categories previously assumed or found (18). We may, even must, change the categories if we want to make progress, enlarge or im-
prove our knowledge, and make our meanings more accurate. But as long as we have the possibility of arranging our information into classes or categories, we have meaning.

Organisation of a system is therefore more than order. And, equally, information is more than entropy (or negentropy). Once we have a system, we have a boundary, hence boundary conditions; these exercise constraints, and with them comes the possibility to build information into the system and give it a structure. We have stored information — redundancy expressing relationship between symbols — and (potential) new information can be abstracted from the system provided it contains sufficient uncertainty. The amount of (potential) information is traditionally measured by its «surprise value». Information is therefore generated if the network of stored information — the redundancy structure — is capable of changing uncertainty into it by imposing rules. Each level of a hierarchical system is given not only by its stored information that has gone into it to produce the level but, equally, also by the new or potential information that it can provide.

I think this approach basically, if not yet in detail, solves the problem of the hierarchy of levels within an informational system. This hierarchy is, however, very different from the causal structure of energy levels within a physical system. Take as example a system made up of gas molecules that can be liquefied or solidified under suitable conditions. A liquid is more organised than a gas, we would say, and the solid represents a still higher level of organisation. But we cannot ascribe a large amount of (potential) information to either of them. The reason is, I think, that the van der Waals forces responsible for the liquid level and even the ordinary valence forces producing the solid are very short range; thus only small patterns of organised material are formed, for example unit cells in the solid. There are no long range effects between the molecules, only repetition; the redundancy is of a minor kind, so is the stored information, and in spite of relatively large random movements or of uncertainty little (potential) information can be extracted from the system.

Think of the human being as such a hierarchy of levels, and the situation is very different. Take as the «lowest» level that of atoms and molecules, the somewhat artificial level of chemical compounds; then continue with the macromolecular level of genes and chromosomes; the cellular, physiological, anatomical and, finally, the mental-emotional levels follow. (There may be more levels, even a continuity of them; this is not under discussion at the moment). We cannot doubt that the so-called higher levels of organisation contain more stored information as well as are capable of giving up more (potential) information. We are convinced that the higher levels, all the same, cannot possibly violate the laws established
on the lower levels. No one can beat the laws of physics, say, by simply willing to escape gravity and fly. We also accept that the higher levels of organisation somehow arise from the lower levels, although they possess more «degrees of freedom». The constraints imposed on higher levels, or the boundary conditions, introduce long-range interrelationships that allow Markov processes of very high order to occur. The high degree of organisation is accompanied by a large amount of uncertainty made possible by the vast number of possible transitions between the states due to stored information. Clearly, we must consider the fact that evolution has brought about this hierarchy of levels.

Evolution is both growth and differentiation. The dinosaur died out because of too much growth (and organisation) and not enough differentiation. Insects have kept their lowly status because they became too specialized so that they could survive only through the «artificial» growth of colonies like the ant hill. Freedom as result of organisation characterizes evolution as an information process.

The most recent measurements of the DNA sequence of cells of various organisms have been described by Gatlin in the following way. «Vertebrates have achieved their high $R$ values by holding $D_1$ relatively constant and increasing $D_2$, whereas the lower organisms (which achieve $R$ values in the vertebrate range and higher) do so primarily by increasing $D_1$. The mechanism is fundamentally different». The amount of $D_2$ is an index for the direction which evolution has taken, «time's arrow» ($^{10}$).

An organism’s hierarchy of levels represents stages of stability and integration. Sherrington already spoke of integrative levels. This integration seems to be brought about by the ever increasing dependence between symbols in the stored information, or the increasing Markov order of redundancy. A thermodynamic system, which is more or less closed against the environment, exhibits a tendency towards the equilibrium state, for the given energy and subject to the boundary conditions. This state is characterized by the minimum value of the free energy. It is tempting to speculate if we can describe the behaviour of an informational system in a similar way. If we identify internal energy with stored information, or redundancy, as I have done previously, noise must play the role of bound energy; but the potential information of the system must then be, it seems, the sum of stored information and noise, though the former must always be greater than the latter if the system is to be viable. Instead of $F = U - TS$, we have $potential I = R + N$, and $R > N$. Should we maximize $potential I$ or minimize $(R - N)$ in order to characterize a level of an informational system?

This is too simple an analogy, I think. Whenever a certain degree of stability or integration is reached on one level, a new level may arise.
subject to constraints due to both the environment and the earlier level, or levels. As long as we see evolution, in the most general way, as the tendency to integration, the finite amount of information available will be organised on one level. Once this level has become stable, the information stored in it can change some of the noise present to provide new relations between symbols or higher-order Markov dependencies that can form another, higher level. The constraints imposed by the boundary conditions are essential for this transformation to occur. The boundary conditions reflect the environment's action but, to an ever increasing degree, also the organisation of the system since the higher levels of an organism become progressively autonomous. It is certainly exemplified by the hierarchy of levels in man. The macro-molecular level of genes displays more interaction between molecules, and forms bigger units, than are known on the atomic or chemical level. The cell, by virtue of its larger unit and greater complexity, is more organised than the gene or chromosome, etc. Interaction — in contrast to one-sided action — can create novel modes of behaviour, over and above the possibilities provided by single action. This is particularly so with non-linear interaction which surely must be predominant in living organisms. A physical illustration is, again, given by the vibrations of a non-linear oscillator that produce the «jump phenomenon», a new mode unknown in the linear case.

The boundary conditions must differ for each level of organisation: and they represent the evolutionary changes the system has undergone in the past and set the stage for new changes if the evolution of the system can continue. The integrative tendency — the drive to form larger wholes — leads to higher levels of organisation but also to more degrees of freedom of the system as a whole. There is not merely addition, or growth, but differentiation as well: how do the constraints imposed by the boundary conditions work? The kind of conditions as we know them from physics — holonome, skleronome, rheonome — cannot easily be imagined to work. According to Gatlin, game theory can prescribe the boundary conditions that are required for the evolution of higher forms (19). A game is any conflict situation between two or more opposing sets of interest. Maxmin and minmax represent the two optimal strategies. Minmax \( D \), and maxmin \( D \) — or maxmin \( H \) and minmax \( H \) — are supposed to play the role of boundary conditions for DNA. An optimization principle must always intervene in any conflict situation so as to prevent extremes. When higher organisms emerged, they did so by placing game theoretic boundaries on the entropy variables of their DNA to minimise the maximum error and maximise the minimum variety... they have selected a successful strategy which advanced them to a higher level in the hierarchy of living systems.
7. — There can be no doubt that molecular biology allows us to construct a model for information transmission greatly superior to that given by electromagnetic theory. The DNA sequence is a highly organised information source and also gives a model for a hierarchical system of levels. Replication, transcription and translation show the details of the transmission process which can be specified even more when initiation, termination, mutation, self-repair, etc. are considered. Finally, the protein synthesis which is the end result of the transmission process can be analysed in terms of pattern recognition. Information is distributed over two, separate semantic dimensions that allow meaning to be specified openly while it is presupposed, or lumped in with, the message in the telegraph model.

Information is more than entropy (or negentropy if you like). The organisation of a system is more than the linear order of a sequence. There emerges a new semantic dimension when stored information in the source, translation in the transmission, and pattern recognition in the reception are needed to produce the message. The stored information, or redundancy, represents a classification, or scheme, for the potential information; it functions as a meta-language in terms of which we can «evaluate», or give meaning to, new information. The translation goes via a dictionary and so involves necessarily a meta-level. The pattern recognition, again, is a classification process through which symbols are arranged into previously established or assumed categories of meaning.

The genetic information model of a natural process is therefore much richer than the telegraph model or the causal model which has only energy transmission. The highly organised systems and the complex interactions between them that we want to explain in the life sciences could never be accommodated satisfactorily in the causal model. Even the telegraph model is not very illuminating, and this is the reason why eminent microbiologists, like Monod, have little regard for information theory. A good illustration is found in the history of western thought which has been plagued for two thousand years by the twin problems of Mind-Body and Free-Will. The model of Man underlying both is that of causality. Thus we are forced to see the mind-body relation as a causal encounter between two «substances»; and our actions must either obey, without fail, the causal law or be capricious and inexplicable.

The causal law, however much we may have succeeded in refining it — from the original Greek law of revenge, through Newtonian determinism to its statistical form in quantum mechanics — is a relic of theology, arising from the moralistic view of the universe. If we accept the history of ideas as a process of mental-emotional evolution of mankind, we can easily see that the progressive elimination of metaphysics is a slow approach to a rational and more realistic view. Thus it is natural that thinking
should have begun with the ideas of an inexorable law, of the certainty of knowledge, and of its absolute objectivity that banished the human element and, with it, meaning from science. Gradually, this simple, narrow and rigid framework of concepts expressing infantile thought gave way to the ideas that the natural law is a scheme imposed by the scientist to allow inferences to be drawn, that knowledge is relative and that it is basically uncertain. The development of modern physical theories exemplifies this progress of ideas. The Uncertainty Principle is the turning point.

If there were no uncertainty, there would be no information. Uncertainty is not simply the reflection of the statistical nature of microphysical phenomena. Uncertainty is a basic fact when knowledge is derived from experimentation. Interaction between measuring instrument and phenomenon must occur and this introduces a conflict between opposing, though equally needed, factors or tendencies. A balance has to be struck by the experimenter between object and subject, phenomenon and instrument, between freedom and constraint, or between new information and background noise. Measurement becomes part of the natural process to be investigated, and information so arises through an organised system which subsumes the single, causal sequence of events. Theory and practice can no longer be absolutely separated, and science as an activity in the laboratory implies that the experimenter's interpretation of the results is essential. Thus the human element and meaning return to the scientific scene from which they had been banned through the excessive idealization of previous, classical theories.

It is natural that we can expect a solution of a human problem like that of mind-body or free-will only if we have a conceptual framework capable of dealing with science as a human activity. This is given by the genetic model of information. If we see ourselves as persons, as systems organised through evolution and experience to a degree that makes us capable to respond to meaning, then we can reach a new understanding of both the external world and ourselves.

REFERENCES

(7) Quastler, H. The emergence of biological organization, University Press, Yale (1964).


(12) Somenzi, V. La filosofia degli automi, Boringhieri, Torino (1965).


